European University Institute

DEPARTMENT OF ECONOMICS

Essays in the Econometrics of Macroeconomic Models

Andreas Tryphonides

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 30 September 2016

European University Institute
Department of Economics

Essays in the Econometrics of Macroeconomic Models

Andreas Tryphonides

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Fabio Canova, EUI, Supervisor
Prof. Peter Hansen, The University of North Carolina at Chapel Hill
Prof. Giuseppe Ragusa, LUISS
Prof. Frank Schorfheide, University of Pennsylvania

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, Andreas Tryphonides, certify that I am the author of the work "Essays in the Econometrics of Macroeconomic Models" I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Signature and Date:

23/06/2016

Μα κι εκεί ο αγώνας του δεν τελείωσε· απάνω στο Σταυρό τον περίμενε ο Πειρασμός [...] το πλανερό δράμα μιας γαλήνιας, ευτυχισμένης ζωης: είχε πάρει, λέει, ὲτσι του φάνηκε, τον εύκολο δρόμο του ανθρώπου [...] θυμόταν τις λαχτάρες της νιότης του και χαμογελούσε ευχαριστημένος, τί καλά, τί φρόνιμα που ὲκαμε [...], και τι παραφροσύνη ὴταν εκείνη να θέλει, λέει, να σώσει τον κόσμο! [...] είπε ὸχι , ὸχι , δεν πρόδωκε, δόξα σοι ο Θεός, δεν λιποτάχτησε, εξετέλεσε την αποστολή του [...] Τετέλεσται!

Νίκος Καζαντζάκης (1955), Ο Τελευταίος Πειρασμός
Nikos Kazantzakis, The Last Temptation [of Christ], Prologue

Abstract

The thesis has focused on issues related to the use of external information in the identification, estimation and evaluation of Dynamic Stochastic General Equilibrium (DSGE) models, and comprises three papers. The first paper, entitled *Improving Inference for Dynamic Economies with Frictions - The role of Qualitative Survey data*, proposes a new inferential methodology that is robust to misspecification of the mechanism generating frictions in a dynamic stochastic economy. I derive a characterization of the model economy that provides identifying restrictions on the solution of the model that are consistent with a variety of mechanisms. I show how qualitative survey data can be linked to the expectations of agents and how this link generates an additional informative set of identifying restrictions. Moreover, I show how the framework can be used to formally validate mechanisms that generate frictions. Finally, I apply the methodology to estimate the distortions in the Spanish economy due to financial frictions and derive an optimal robust Taylor rule. The second chapter, entitled *Estimation and Inference for Incomplete Structural Models using Auxiliary Density Information* considers an alternative method for estimating the parameters of an equilibrium model which does not require the equilibrium decision rules and produces an estimated probability model for the observables. This is done by introducing auxiliary information about the conditional density of the observables, and using density projections. I develop and assess frequentist inference in this framework. I provide the asymptotic theory for parameter estimates for a general set of conditional projection densities and simulation exercises. In the third chapter, entitled *Monetary Policy Rules and External Information*, I analyze how conclusions about monetary policy stance are altered when we explicitly acknowledge that model concepts like the output gap and inflation are non-observable and we utilize many proxies that are available in the data. I document the effects on Bayesian inference of introducing such proxy information.

Acknowledgements

Firstly, I would like to express enormous gratitude and appreciation to my advisor, Fabio Canova, for many reasons. Fabio has not been just a thesis advisor but a great mentor and I would like to thank him for his support, generosity and patience and for believing in me. His help and guidance was crucial in many steps of writing the thesis, especially Chapter 2. Within five years, I have learned a lot from him, and the mere thesis content reflects maybe a fraction of it. I hope I will manage to become a teacher and mentor as exceptional as he has been to me.

I would also like to thank Peter Hansen for his teaching and occasional criticism. Same goes to Juan Jose Dolado, who was always willing to hear what I had to say and give advice. I am also very thankful to Evi Pappa for her help and advice during the job market. This thesis has also benefited from teaching and interactions I had with members of faculty during my visit at the University of Wisconsin Madison, and from various comments and discussions in conferences and seminars I have participated. In addition, I would like to thank the thesis examiners Frank Schorfheide and Giuseppe Ragusa for useful comments and suggestions.

Moreover, I acknowledge the financial support for my studies from the Cyprus Ministry of Education and the EUI.

During these five years at the EUI, I have met (or reunited with) a lot of great people with which I have interacted in various ways, and I am sure that some of them will stay close for many years to come. I can't help not mentioning Michalis, Romanos, John, Brais, Nacho and Valentinos.

Special thanks goes to all the staff at Villa San Paolo, and especially Jessica Spataro and Lucia Vigna who were always there to deal with all my requests and needs.

I would like to thank Lena for her support and patience during the last months of finishing the thesis. Lastly, I would like to express my gratitude to my parents, for making all this possible and my sister, Alexia, for being the only constant in a sea of change; this thesis is dedicated to her.

Contents

List of Tables

List of Figures

Chapter 1

Summary

The thesis has focused on issues related to the use of external information in the identification, estimation and evaluation of Dynamic Stochastic General Equilibrium (DSGE) models, and comprises three papers.

The first paper, entitled **"Improving Inference for Dynamic Economies with Frictions - The role of Qualitative Survey data"**, proposes a new inferential methodology that is robust to misspecification of the mechanism generating frictions in a dynamic stochastic economy. The motivation comes from the fact that while frictions present in Dynamic Stochastic General Equilibrium (DSGE) models are essential to match the persistence observed in aggregate macroeconomic time series, their selection and specification is a complicated process and features arbitrary aspects. Thus, different studies may find support for different types of frictions depending on auxiliary modeling assumptions. Moreover, unless frictions are micro-founded, policy conclusions are whimsical as different mechanisms imply different relationships between policy parameters and economic outcomes.

The proposed approach treats economies with frictions as perturbations of a frictionless economy. Due to the fact that a variety of mechanisms are consistent with the same model representation, models and their parameters are set identified. In order to reduce model uncertainty, additional restrictions coming from qualitative survey data are utilized, and it is shown that, under certain conditions, they can help to disentangle otherwise observationally equivalent models. The econometric theory developed can accommodate conditional moment restrictions constructed from data other than surveys.

The paper makes three contributions. First, it presents a characterization of the model economy that provides identifying restrictions on the solution of the model. It is shown that the sign of the conditional mean of the distortion induced by the frictions is known, even if

the exact mechanism generating the friction is unknown. If the decision rules of the economy with frictions are: $X_t = X_t^f + \lambda_t$ where X_t^f is the solution of the frictionless model and λ_t an unobserved variable, one can use equations of the form $\mathbb{E}_t \lambda_t \geq 0$ or $\mathbb{E}_t \lambda_t \leq 0$, depending on the sign of conditional mean distortion to set-estimate the parameters.

Second, it shows how qualitative survey data can be linked to the expectations of agents and how this link generates an additional set of identifying restrictions on the probability of observing a distortion (relative to the frictionless model) in a macroeconomic variable. When agents report their subjective conditional expectation, and under certain conditions, the additional restrictions lead to a smaller set of admissible models.

Third, it shows how the framework can be used to validate mechanisms that generate frictions. It develops a test statistic that examines the minimum distance of the point identified complete model to the admissible set of models. Under the null, that the fully parametrically specified model belongs to the identified set of models, the expected minimum distance is zero. Under the alternative, it is strictly positive. Large sample theory for this test is derived and a bootstrap procedure to compute the critical values is proposed. The test is asymptotically powerful against fixed alternatives. A Monte Carlo exercise confirms the asymptotic results.

Finally, the methodology is applied to estimate the distortions in the Spanish economy due to financial frictions. A small open economy version of the (Smets and Wouters 2007) model is employed and qualitative survey data on the financial constraints of the agents, collected by the European Commission, is utilized. I identify the model generating the maximum distortion to observed macroeconomic variables that is consistent with survey data and derive the corresponding optimal Taylor rule.

The second paper, entitled **"Estimation and Inference for Incomplete Structural Models using Auxiliary Density Information"**, deals with the fact that while macroeconomic theory provides a set of equilibrium moment conditions, it rarely provides the complete probability distribution of observables, 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. These choices may induce misspecification and loss of identification power. On the other hand, having a complete probability distribution is useful, since one can make counterfactual experiments and predictions.

This paper considers an alternative method for estimating the parameters of a DSGE model which does not require the equilibrium decision rules and produces an estimated probability model for the observables. The model is completed by introducing information about the conditional density of the observables $f(x|z,\phi)$, where z is the relevant conditioning information. This density can be generally interpreted as an approximate reduced form model for the observables. Utilizing a variation of the method of information projections, see for example (I.Csiszar 1975) for the unconditional case, I obtain the conditional probability distribution that satisfies the moment restrictions provided by economic theory, $\mathbb{E}(m(x,\vartheta)|z) = 0$, and is as close as possible to the prior density. This is also related to the work of (Giacomini and Ragusa 2014) where density projections are used in a forecasting context.

The methodology can allude to the Bayesian paradigm in the sense that the approximate model serves as a prior, which can nevertheless be data revisable. The paper also provides a decision theoretic framework that rationalizes the estimator as an optimal outcome of a two stage Stackelberg game between the leader, the Principal, who constrains the set of probability distributions considered using prior information, and then delegates parameter estimation to the follower, the Econometrician.

The paper mainly develops and assesses frequentist inference in this framework. It provides the asymptotic theory for parameter estimates for a general set of conditional projection densities. Under the condition that there exists an admissible parameter of $f(x|z,\phi)$ such that the moment conditions are satisfied, the estimator is consistent and asymptotically Normally distributed, where the semi-parametric lower bound for the parameter estimates, is attained. More interestingly, in the case of density misspecification, the first order conditions for estimating structural parameters are akin to penalizing the standard optimally weighted Generalized Method of Moments (GMM) estimator with a penalty term that goes to zero as the approximate density becomes closer, in the total variation distance, to the true density. Under local misspecification of the density in the form of improper finite dimensional restrictions, there exist efficiency gains and therefore an asymptotic bias - variance trade-off. Monte Carlo simulations are performed for the unconditional moment case, and comparisons of the Mean Squared Error (MSE) of the estimator are provided in the case of fixed misspecification. More exercise are done using the prototypical stochastic growth model.

The third paper, entitled **"Monetary Policy Rules and External Information"**, analyzes how conclusions about monetary policy stance are altered when I explicitly acknowledge that model concepts like the output gap, inflation, and nominal interest rate are non-observable. On the other hand, many proxies are available in the data. For example, there exist a number of measures of inflation, several types of interest rates and other inflation expectation measures obtained form surveys, nowcasts and professional forecasts. I assume that the vector of observables is a noisy measure of model variables and perform Bayesian estimation of a textbook New Keynesian model augmented with the vector of proxies, see for example (Del Negro and Schorfheide 2013). I document a significant change in the posterior distributions of all the parameters relative to the case when only one proxy measure is employed and no allowance is made for unobservable model concepts. Posterior distributions are not only less dispersed but have different location and support. It is shown that under a measurement error interpretation, the additional information improves the estimates of the state variables of the model and this explains the lower variance of the posterior distributions. The significant change in the location indicates that the standard approach faces both identification and misspecification problems.

Chapter 2

Improving Inference for Dynamic Economies with Frictions - The role of Qualitative Survey data

2.1 Introduction

Frictions are essential in Dynamic Stochastic General Equilibrium (DSGE) models to match the persistence observed in aggregate macroeconomic time series. However, the selection between real, nominal and informational frictions and specification of the functional forms is a complicated process featuring arbitrary aspects. This arbitrariness means that different studies may find support for different types of frictions, depending on other assumptions made. For example, the choice of nominal rigidities might depend on which real rigidities are included in the model or whether there is variable capital utilization or not[1]. Moreover, unless frictions are micro-founded, policy conclusions may become whimsical as different mechanisms imply different relationships between policy parameters and economic outcomes.

In this paper we propose a new inferential methodology that is robust to misspecification of the mechanism generating frictions. The approach treats economies with frictions as perturbations of an ideal frictionless economy. These perturbations are not uniquely pinned down since different specifications of functional forms may lead to observationally equivalent perturbations. For example, a perturbation in the law of motion of capital can be consistent with many exogenous and endogenous capital adjustment costs specifications.

The approach we propose consists of three steps. First, we derive a characterization of the economy that imposes identifying restrictions on the solution of the model. (Chari, Kehoe, and McGrattan 2007), CKM hereafter, identify wedges in the optimality conditions of a frictionless model that produce the same equilibrium allocations in economies with specific parametric choices for the frictions. We also take a frictionless model as a benchmark but contrary to CKM we construct a general representation for the wedge in the decision rules. We illustrate through examples that the sign of the conditional mean of the decision rule wedge is typically known, even when the exact mechanism generating the friction is unknown. We provide necessary and sufficient conditions such that the decision rules of linear (or linearized) economies with frictions can be represented as $X_t = X_t^f + \lambda_t$, where X_t^f is the solution of the frictionless model and λ_t a latent process. If these conditions are satisfied, one can use equations of the form $\mathbb{E}_t \lambda_t \geq 0$ or $\mathbb{E}_t \lambda_t \leq 0$, depending on the sign of the conditional mean of the wedge, to set-identify the parameters of the model. Moment inequality restrictions have been used to characterize frictions in specific markets, see for example (Luttmer 1996) and (Chetty 2012). We are the first to apply

[1] See for example the analysis of (Christiano, Eichenbaum, and Evans 2005) and their comparison to the results of (Chari, Kehoe, and McGrattan 2000).

the methodology to dynamic stochastic macroeconomic models and to show their relationship with the wedge literature. We deal with partial identification that moment inequalities imply by employing recent advances on bounds estimation (i.e. (Liao and Jiang 2010, Chernozhukov, Kocatulum, and Menzel 2012)).

Because we partially identify economic relationships to avoid misspecification, a large set of models is likely to be consistent with the same identification assumptions. Thus, additional data is needed to separate otherwise observationally equivalent models. We use qualitative survey data for this purpose.

In the second step, we link qualitative survey data to the expectations of agents. If survey data reflect subjective conditional expectations, they contain important indications about agents' actions and thus provide information on which model may be inconsistent with the data. Therefore, the set of admissible models may be reduced. In the literature [2] survey data are typically linked to an observable model, and an explicit mapping between theory based and observable variables is made with an additive measurement error (see e.g. (Del Negro and Schorfheide 2013)). Such an approach however cannot be used with qualitative survey data due to their categorical form.

Because the representation we derive accommodates both moment equalities and inequalities, we can link the solution of the model to survey data of different types. In the case of categorical data, we show that they imply restrictions on the probability of observing a friction (relative to the frictionless model) and that, under certain conditions, the additional restrictions lead to a smaller set of admissible models. The econometric theory we develop is sufficiently general and accommodates conditional moment restrictions obtained from data other than surveys.

As a third step, we show how the framework can be used to validate particular specifications generating frictions. We exploit the fact that, once qualitative survey data are used, we have a smaller set of models consistent with the same economic conditions. We propose a test statistic that examines the minimum distance of the point identified parametrically specified model to the admissible set of models. Under the null, that the parametrically specified model belongs to the identified set of models, the expected minimum distance is zero. Under the alternative, it is strictly positive. We derive large sample theory for this test and propose a bootstrap procedure

[2] We mainly refer to the treatment of survey data in the most recent "modern" DSGE literature. The treatment of survey data in Rational Expectations models date back to the work of (Pesaran 1987) and others.

to compute the critical values. We show that the test is asymptotically powerful against fixed alternatives. A Monte Carlo exercise confirms the asymptotic results.

We apply the methodology to estimate the distortions present in the Spanish economy due to financial frictions. We estimate a small open economy version of the (Smets and Wouters 2007) model using qualitative survey data collected by the European Commission on the financial constraints of the agents. We identify the model generating the maximum distortion that is consistent with survey data and derive the corresponding optimal Taylor rule.

The paper is linked to four different strands of literature: the literature on wedges and frictions (i.e. (Chari, Kehoe, and McGrattan 2007), (Christiano, Eichenbaum, and Evans 2005), (Chari, Kehoe, and McGrattan 2000)), the partial identification literature (i.e. (Chernozhukov, Lee, and Rosen 2013), (Chernozhukov, Kocatulum, and Menzel 2012), (Chernozhukov, Hong, and Tamer 2007), (Liao and Jiang 2010)), the literature on robustness to model misspecification (i.e. (Hansen 2013), (Hansen and Sargent 2005)), and the literature on including survey data in DSGE models (i.e. (Del Negro and Schorfheide 2013)). Relative to (Chari, Kehoe, and McGrattan 2007), we adopt a more general and different characterization of the wedge in equilibrium conditions. This general characterization is what makes our approach robust to misspecification, as it accommodates different plausible mechanisms. Moreover, we link the characterization of the wedge in equilibrium conditions to an equivalent characterization for the law of motion of macroeconomic variables. This is a very useful result as the law of motion, compared to equilibrium conditions, can be easily linked to the data. As far as partial identification is concerned, we utilize the quasi Bayesian MCMC methodology of (Liao and Jiang 2010) with uniform priors. We deal with multiple moment inequalities by adopting a single inequality approximation (see (Chernozhukov, Kocatulum, and Menzel 2012)). We also contribute to the literature that deals with partial identification in structural macroeconomic models (e.g. (**?**)). Although we use moment inequalities to represent model uncertainty, we also provide an interpretation of decision rules wedges as causing a change on the probability measure of the frictionless model. Following the semi-parametric approach in (Hansen 2013), we provide a characterization of frictions over time and not just on average. Finally, relative to (Del Negro and Schorfheide 2013), we use inequality restrictions which can accommodate both deviations of the model prediction from aggregate macroeconomic data and qualitative survey data. Apart from enabling us to accommodate qualitative survey data, inequality restrictions on the observation equation provide a more general way of linking models' predictions to the data. This is not only true for DSGE models, but also any model that has a state space representation.

The rest of the paper is organized as follows. Section 2 provides a motivating prototype economy which can be used as an experimental lab for the econometric method that we suggest. Section 3 examines the distortions present in the decision rules and their observable aggregate implications. Section 4 introduces qualitative survey data and derives the relevant bounds. Section 5 provides necessary and sufficient conditions for the correct specification of the moment conditions constructed from survey data, and identification results. Section 6 illustrates how the identified set can be used to test models with specific frictions. Section 7 has an application to Spanish data. Section 8 concludes and provides avenues for future extensions. Appendix A contains the proofs, a simulation result for the proposed test statistic, a discussion on sufficient conditions for estimating the identified set using Markov Chain Monte Carlo (MCMC). Appendix B contains computational results for the case of capital adjustment costs, some plots of survey time series and further details of the data in the application.

Throughout the paper we refer to three different probability measures. The objective probability measure, $\mathbb{P}_t$, the probability measure determined by the frictionless model M_f, $\mathbb{P}_t(.|M_f,.)$, and the subjective probability measure $\mathcal{P}_{t,i}$ where i identifies agent i and t the timing of the conditioning set. All three are absolutely continuous with respect to Lebesgue measure. The corresponding conditional expectation operators are $\mathbb{E}_t(.)$, $\mathbb{E}_t(.|M_f,.)$ and $\mathcal{E}_{i,t}(.)$. We distinguish between the first and the second as the model will be correctly specified only if the econometrician has the right DGP. T denotes the length of both the aggregate and average survey data, and L the number of agents. We denote by $\theta \in \Theta$ the parameters of interest, and by $q_j(;\theta)$ a measurable function. Bold capital letters e.g. $\mathbf{Y}_t$ denote a vector of length t containing $\{Y_j\}_{j\leq t}$. The operator $\rightarrow_p$ indicates convergence in probability and the operator $\rightarrow_d$ convergence in distribution; $\mathcal{N}(.,.)$ is the Normal distribution; $||.||$ is the Euclidean norm unless otherwise stated; $\perp$ indicates the orthogonal complement and $\emptyset$ the empty set. We denote by $(\Omega, \mathcal{S}_y, \mathbb{P})$ the probability triple for the observables to the econometrician, where $\mathcal{S}_y = \sigma(Y)$, is the sigma-field generated by Y and $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, \mathbb{P})$ the corresponding filtered probability space.

2.2 A Working Example

This section considers a simple frictionless economy and illustrates what are the observable implications of certain types of frictions. We call frictions mechanisms creating a wedge in the equilibrium conditions of the frictionless model. Thus, frictions produce differences between the expectations of the variable, conditional on the frictionless model and the available information set, and the data.

Consider a simple Real Business Cycle (RBC) model. The economy is populated with a continuum of agents with Constant Relative Risk Aversion (CRRA) utility, each forming expectations about key decision variables and making basic consumption - savings decisions. They rent capital to a representative firm, which is used for production, and receive a share of profits. Individual investment decisions increase the availability of capital for next period, up to a certain level of depreciation. We denote aggregate variables by capital letters i.e. $X_t \equiv \int x_{i,t} d\Lambda(i,t)$, where $\Lambda(i,t)$ is the distribution of agents at time t. The problem is:

$$\max_{\{c_{i,t}\}_1^\infty} \mathcal{E}_{i,0} \sum_{t=1}^{\infty} \beta^t \frac{c_{i,t}^{1-\omega}}{1-\omega}$$

subject to the following constraints:

$$\begin{aligned} i_{i,t} + c_{i,t} &= \hat{y}_{i,t} = R_t k_{i,t} + pr_{i,t} \\ k_{i,t+1} &= (1-\delta)k_{i,t} + i_{i,t} \end{aligned}$$

where $pr_{i,t} = \eta_{i,t}(Z_t K_t^\alpha - R_t K_t)$, for $\alpha > 0$.

The system of linearized aggregate equilibrium conditions, where ξ_t is the aggregation residual, and $\tilde{X}$ represents deviations of any aggregate variable X_t from aggregate steady state x_{ss} is:

$$\begin{aligned} -\omega \tilde{C}_t - \xi_{C,t} - \beta r_{ss} \bar{\mathcal{E}}_t \tilde{R}_{t+1} + \omega \bar{\mathcal{E}}_t \tilde{C}_{t,+1} &= 0 \\ y_{ss}\tilde{Y}_t - c_{ss}\tilde{C}_t - i_{ss}\tilde{I}_t &= 0 \\ \tilde{K}_{t+1} - (1-\delta)\tilde{K}_t - \delta \tilde{I}_t &= 0 \\ \tilde{Y}_t - \tilde{Z}_t - \alpha \tilde{K}_t - \xi_{Y,t} &= 0 \\ \tilde{R}_t - \tilde{Z}_t + (1-\alpha)\tilde{K}_t - \xi_{R,t} &= 0 \end{aligned}$$

Under approximate linearity, that is, when ξ_t is negligible [3], we can easily obtain the aggregate decision rules for X_t. The equilibrium decision rules will depend on predetermined capital

[3] Approximate linearity has implications beyond the choice of functional form. In particular, we assume complete markets. Nevertheless, note that incomplete markets is a friction, and thus can be one of the deviations one can consider. We discuss this in the last two sections and Appendix B.

$\tilde{k}_{i,t}$ and on expectations of the aggregate productivity shock for j periods ahead, $\mathcal{E}_{i,t}(Z_{t+j}|X_t)$. For example, the aggregate investment decision rule is:

$$\tilde{I}_t = A_1(\theta)\tilde{K}_t + A_2(\theta)\sum_{j=0}^{\infty} A_3(\theta)^j \bar{\mathcal{E}}_t(\tilde{Z}_{t+j}|X_t) \tag{2.1}$$

where $A_j(\theta), j = 1,2,3$ are functions of the structural parameters θ and $\bar{\mathcal{E}}_t$ indicates aggregate conditional expectations. Note that individual expectations $\mathcal{E}_{i,t}$ are not necessarily formed with respect to the objective probability measure, nor with respect to the same information set.

As a benchmark case, consider the case when the econometrician has the same model as the agents, that is the true model has no frictions, and agents have rational expectations i.e. $\mathcal{E}_i(.|X_t) = \mathbb{E}(.|X_t)$. Here, it is typical to assume that only a subset of the information set of the agents is observed by the econometrician, that is $\sigma(X_{1,t}) \subset \sigma(X_t)$. However, since the model is well specified, on average, the difference in the conditional expectations of the econometrician and of the agents is negligible and parameters can be consistently estimated. To see this, notice that the decision rule used by the econometrician can be rewritten as:

$$\tilde{I}_t^{\star} = A_1(\theta)\tilde{K}_t + A_2(\theta)\sum_{j=0}^{\infty} A_3(\theta)^j \mathbb{E}(\tilde{Z}_{t+j}|X_{1t}) + e_t \tag{2.2}$$

where

$$e_t = A_2(\theta)(\sum_{j=0}^{\infty} A_3(\theta)^j \mathbb{E}(\tilde{Z}_{t+j}|X_t) - \sum_{j=0}^{\infty} A_3(\theta)^j \mathbb{E}(\tilde{Z}_{t+j}|X_{1,t}))$$

In this setup, as (Hansen and Sargent 2013) have shown, due to the law of iterated expectations the following moment equalities hold:

$$\mathbb{E}(e_t \phi(X_{1,t-j})) = 0$$

for any $j \geq 1$ and $X_{1,t-j}$ measurable function $\phi(.)$.

The case of interest of this paper is when $\mathcal{E}_i(.|X_{i,t}) \neq \mathbb{E}(.|X_t, M_f)$. Here the agents and the econometrician not only have different information, but they also have a different structure of the economy in mind. We explicitly condition on M_f to emphasize that $\mathbb{E}(.|X_t, M_f)$ is consistent with frictionless behavior and implicitly assume that the frictionless part of the model is well specified. The mismatch between the agents' expectations and the econometrician's prediction could be due to differences in the models and/or information sets. As long as informational differences affect agents' behavior, both differences have similar implications for the decision rules. The individual decision rule is:

$$0 = \mathcal{E}_i(i_{i,t} - A_1(\theta)k_{i,t} + A_2(\theta)\sum_{j=0}^{\infty} A_3(\theta)^j z_{i,t+j}|X_t)$$

and the corresponding aggregate condition is:

$$0 = \int \mathcal{E}_i(i_{i,t} - A(\theta)k_{i,t} + A_2(\theta)\sum_{j=0}^{\infty} A_3(\theta)^j z_{i,t+j}|X_t)d\Lambda(i) \tag{2.3}$$

Real, nominal or informational frictions make (2.3) different from (2.1), and therefore (2.2). As we show below, these differences can be characterized by moment inequalities. We consider three examples: adjustment costs, occasionally binding constraints, and non-rational expectations. These frictions constrain agents' optimal behavior and generate a "wedge" $\tilde{\lambda}_t$ that can be used to construct moment conditions. For i^* the unconstrained optimal choice of investment and i_{con} the constrained optimal choice and aggregating across agents, we have the following representation:

$$\begin{aligned} \tilde{I}_t^\star &= A_1(\theta)\tilde{K}_t + A_2(\theta)\sum_{j=0}^{\infty} A_3(\theta)^j \mathbb{E}(\tilde{Z}_{t+j}|X_{1t}) + e_t + \tilde{\lambda}_t \\ \tilde{\lambda}_t &= I_{con,t} - I_t^* \end{aligned}$$

Example 1. *Capital Adjustment Costs*
Assuming full depreciation, the capital accumulation equation of the representative firm is distorted as follows: $K_{t+1} = I_t - \frac{\phi}{2}(\frac{K_{t+1}}{K_t} - 1)^2 K_t$ *for* $\phi \in (0,1)$.

Using the capital accumulation equation in the linearized Euler equation and imposing $\tilde{R}_t = \tilde{Z}_t - (1-\alpha)\tilde{K}_t$ *we have:*

$$\begin{aligned} (\omega + \phi(1+\beta(1-\alpha)) + 1 - \alpha)\kappa_t &= (\alpha - \phi(1-\beta\alpha))(\alpha-1)\tilde{K}_t + (\omega+\beta\phi)\kappa_{t+1} - \phi\tilde{Z}_t \\ &= (\alpha - \phi(1-\beta\alpha))\tilde{R}_t + (\omega+\beta\phi)\mathbb{E}_t\kappa_{t+1} + \ldots \\ &\quad \ldots - \alpha(1-\beta\phi)\tilde{Z}_t \end{aligned}$$

where κ_t *is the Lagrange multiplier on the capital accumulation equation. Assuming that pro-*

ductivity is iid, and iterating forward we get:

$$\kappa_t = \frac{\gamma_1}{1-\gamma_2 L^{-1}}\mathbb{E}_t\tilde{R}_t + \gamma_3\tilde{Z}_t = \zeta(L,\gamma_1,\gamma_2)\mathbb{E}_t\tilde{R}_t + \gamma_3\tilde{Z}_t$$

where $\gamma_1 \equiv \frac{\alpha-\phi(1-\beta\alpha)}{\omega+\phi(1+\beta(1-\alpha))+1-\alpha}$, $\gamma_2 = \frac{\omega+\beta\phi}{\omega+\phi(1+\beta(1-\alpha))+1-\alpha}$ *and* $\gamma_3 = \frac{-\alpha(1-\beta\phi)}{\omega+\phi(1+\beta(1-\alpha))+1-\alpha}$ *and* L *the lag operator. Using* $\kappa_t = -\omega\tilde{C}_t$, *and letting* $s := \frac{C_{ss}}{I_{ss}}$, *aggregate investment is:*

$$\tilde{I}_{con,t} = (1+s)\tilde{Y}_t + \frac{s}{\omega}\zeta(L,\gamma_1,\gamma_2)\mathbb{E}_t\tilde{R}_t + \frac{s}{\omega}\gamma_3\tilde{Z}_t$$

When $\phi = 0$,

$$\tilde{I}^{\star}_t = (1+s)\tilde{Y}_t + \frac{s}{\omega}\zeta(L,\gamma_{1,\phi=0},\gamma_{2,\phi=0})\mathbb{E}_t\tilde{R}_t + \frac{s}{\omega}\gamma_{3,\phi=0}\tilde{Z}_t \tag{2.4}$$

Notice that $\zeta(L,\gamma_1,\gamma_2)$ *is increasing in both* γ_1 *and* γ_2 *and* $\frac{d\gamma_1}{d\phi} < 0$, $\frac{d\gamma_2}{d\phi} < 0$ *for* $\omega > 1$ *and* $\frac{d\gamma_3}{d\phi} > 0$. *Therefore,* $\zeta(L,\gamma_{1,\phi=0},\gamma_{2,\phi=0}) - \zeta(L,\gamma_1,\gamma_2) > 0, \forall\phi$ *and the difference the investment rules with and without adjustment costs is:*

$$\begin{aligned}
\tilde{\lambda}_t &\equiv \tilde{I}_{con,t} - \tilde{I}^{\star}_t \\
&= \frac{s}{\omega}(\zeta(L,\gamma_{1,\phi=0},\gamma_{2,\phi=0}) - \zeta(L,\gamma_1,\gamma_2))\mathbb{E}_t\tilde{R}_t + \frac{s}{\omega}(\gamma_{3,\phi=0}-\gamma_3)\tilde{Z}_t \\
&= -\frac{s}{\omega}(\zeta(L,\gamma_{1,\phi=0},\gamma_{2,\phi=0}) - \zeta(L,\gamma_1,\gamma_2))(1-\alpha)\tilde{K}_t ... \\
&\quad ... + \frac{s}{\omega}(\zeta(L,\gamma_{1,\phi=0},\gamma_{2,\phi=0}) - \zeta(L,\gamma_1,\gamma_2) + (\gamma_{3,\phi=0}-\gamma_3))\tilde{Z}_t
\end{aligned}$$

Hence, $\tilde{\lambda}_t$ *is negatively related to* $\tilde{K}_t$. *Moreover, after some algebra it can be shown that the coefficient of* $\tilde{Z}_t$ *is also bounded below by a positive number if* $\alpha < \frac{1}{2\beta}$. *Nevertheless, the sign of the conditional mean of* $\tilde{\lambda}_t$ *is determined and the following moment inequality holds:* $\mathbb{E}\tilde{\lambda}_t\tilde{K}_{t-j} \leq 0, \forall j \geq 0$.

Example 2. *Occasionally binding constraints*

The case of occasionally binding constraints can be best motivated by attaching an aggregate marginal efficiency shock, ε_t *to investment, that is,* $k_{i,t} = (1-\delta)k_{i,t-1} + \varepsilon_t i_{i,t}$. *For simplicity we assume that this shock is iid and takes two values,* ε_H *and* ε_L. *We analyze the case of constraints on dis-investing (capital irreversibility), which can be thought of as a restriction on how much capital households can withdraw from the firm every period. The optimization problem now includes a new constraint of the form* $K_t \geq \rho(1-\delta)K_{t-1}$ *which is equivalent to* $I_t \geq -\frac{\tilde{\rho}}{\varepsilon_t}K_t$ *where* $\tilde{\rho} \equiv (1-\rho)(1-\delta)$. *Denoting by* ν_t *the Lagrange multiplier on this constraint, and* κ_t

the Lagrange multiplier on the law of motion of capital, the relevant optimality conditions are:

$$\begin{aligned}
\kappa_t + \beta \mathbb{E}\kappa_{t+1}(\tilde{\rho} - (1-\delta)) + \mathbb{E}C_{t+1}^{-\omega}(R_{t+1} + \frac{\tilde{\rho}}{\varepsilon_{t+1}}) &= 0 \\
\nu_t - \kappa_t \varepsilon_t - C_t^{-\omega} &= 0 \\
\nu_t(I_t + \frac{\tilde{\rho}}{\varepsilon_t}K_t) &= 0
\end{aligned}$$

When $\varepsilon_t = \varepsilon_H$ the representative household will choose $I^\star_{H,t}$ according to the following Euler equation, which we get by setting $\nu_t = 0$:

$$C_t^{-\omega} = \mathbb{E}\beta C_{t+1}^{-\omega}\frac{\varepsilon_{H,t}}{\varepsilon_{t+1}}(1 - \delta + R_{t+1}\varepsilon_{t+1})$$

and linearizing we have:

$$-\omega\tilde{C}_t = -\mathbb{E}\tilde{C}_{t+1} + \tilde{\varepsilon}_{H,t} + \mathbb{E}\tilde{R}_{t+1}$$

Solving the Euler equation forward, we get $\tilde{I}^\star_{H,t} = \tilde{I}^\star_t + \tilde{\varepsilon}_{H,t}$. When $\varepsilon_t = \varepsilon_{L,t}$, the household disinvests up to the irreversibility level, that is $I_{L,t} = -\frac{\tilde{\rho}}{\varepsilon_t}K_t$. The corresponding linearized rule is $\tilde{I}_{L,t} = -(\tilde{K}_t - \tilde{\varepsilon}_t)$ where we have imposed that $\mathbb{E}\varepsilon_t = 1$. Therefore aggregate investment evolves as:

$$\begin{aligned}
\tilde{I}_{con,t} &= \tilde{I}^\star_{H,t}\mathbb{P}(\varepsilon_t = \varepsilon_H) + I_{L,t}\mathbb{P}(\varepsilon_t = \varepsilon_L) \\
&= \tilde{I}^\star_t - (1 - \mathbb{P}(\varepsilon_t = \varepsilon_H))(\tilde{I}^\star_{H,t} - \tilde{I}_{L,t})
\end{aligned}$$

and the distortion produced by the occasionally binding constraint is:

$$\tilde{\lambda}_t = \tilde{I}_{con,t} - \tilde{I}^\star_t = -(1 - \mathbb{P}(\varepsilon_t = \varepsilon_H))(\tilde{I}^\star_t + \tilde{K}_t - \tilde{\varepsilon}_L + \tilde{\varepsilon}_H)$$

Here $\tilde{\lambda}_t$ is negative as by definition $\tilde{\varepsilon}_H > \tilde{\varepsilon}_L$. The moment inequality implied by this friction is $\mathbb{E}\tilde{\lambda}_t\tilde{K}_{t-j} \leq 0, \quad \forall j \geq 0$.

Example 3. *Non Rational Expectations*

When agents employ different models to make predictions, have mis-perceptions or sentiments, the sign of $\mathbb{E}\tilde{\lambda}_t\tilde{K}_{i,t-j}$ depends on how the model used by the agent relates to the objective probability measure. Suppose, for illustration, that agents are unaware and unable to estimate the stochastic process for productivity. Suppose that the true process is $\tilde{Z}_t = \varepsilon_t$ where $\varepsilon_t \sim N(0,1)$.

Agents use output realizations to predict future productivity, $\mathcal{E}_t\tilde{Z}_{t+j} = \rho^j\tilde{Y}_t$ *for* $|\rho| < 1$*, since* $Corr(Z_t, Y_t) = \frac{1}{1+\alpha^2\mathbb{V}(K_t)} > 0$*. Using this conditional expectation in the investment rule 2.1, aggregate investment is:*

$$\tilde{I}_{con,t} = A_1(\theta)\tilde{K}_t + A_2(\theta)(1 - A_3(\theta)\rho)^{-1}(\alpha\tilde{K}_t + \tilde{Z}_t)$$

while the investment rule used by the econometrician after substituting the true process for $\tilde{Z}_t$ *in* (2.1) *is:*

$$\tilde{I}_t = A_1(\theta)\tilde{K}_t + A_2(\theta)\tilde{Z}_t \tag{2.5}$$

The difference between the investment rule under bounded rationality and the one used by the econometrician is therefore:

$$\tilde{\lambda}_t = A_2(\theta)(1 - A_3(\theta)\rho)^{-1}(\alpha\tilde{K}_t + \rho A_3(\theta)\,\tilde{Z}_t)$$

Assuming full capital depreciation and using the true process for productivity in equation (2.5)*, which is identical to equation* (2.4) *once we substitute for the production function and the return to capital, we get that* $A_2(\theta) = 1 + \frac{C_{ss}}{I_{ss}} > 0$ *and* $A_3(\theta) = \frac{s\alpha}{(1+s)(\omega+1-\alpha)} < 1$*. Therefore,* $\tilde{\lambda}_t$ *is a positive function of* $\tilde{K}_t$ *and* $\tilde{Z}_t$*, and the moment inequality in this case is,* $\mathbb{E}\tilde{\lambda}_t K_{t-j} \geq 0, \quad \forall j \geq 0$.

In all of the examples, $\tilde{\lambda}_t$ does not behave as classical measurement error and is endogenous, as it depends on capital. Thus, the setup is similar to the endogeneity problem in standard regressions where the omitted information is correlated with the regressors. By imposing the inequalities we acknowledge this endogeneity.

The examples deal with distortions in the decision rules. In the next section, we show how to translate distortions to the equilibrium conditions into observationally equivalent distortions to decision rules. Distortions to decision rules are economically more informative as they directly affect the transmission of shocks and welfare and provide a framework where survey data can be introduced.

Because we will work with linearized models, second or higher order effects will be ignored. However, moment inequalities, would also appear in nonlinear models. We also ignore the presence of approximation errors because the way we treat frictions does not depend on assumptions regarding the approximation error.

2.3 Perturbing the Frictionless Model

Denote by $x_{i,t}$ the endogenous individual state, by $z_{i,t}$ the exogenous individual state, and by $X_t = \int x_{i,t} d\Lambda(i)$ and $Z_t = \int z_{i,t} d\Lambda(i)$ the corresponding aggregate states.

Assume that the optimality conditions characterizing the individual decisions are:

$$\begin{aligned} G(\theta)x_{i,t} &= F(\theta)\mathcal{E}_{i,t}\left(\begin{pmatrix} x_{i,t+1} \\ X_{t+1} \end{pmatrix} | x_{i,t}, z_{i,t}, X_t, Z_t\right) + L(\theta)z_{i,t} \qquad (2.6) \\ z_{i,t} &= R(\theta)z_{i,t-1} + \varepsilon_{i,t} \end{aligned}$$

where $\mathbb{E}(\varepsilon_{i,t}) = 0$. Because we assume that the coefficients of the optimality conditions are common across agents, preferences and technologies are common across agents. Relaxing this assumption would make the notation more complicated, but would not change the essence of the argument. We could also specify equilibrium conditions that involve past endogenous variables but this is unnecessary as we can always define dummy variables of the form $\tilde{x}_{i,t} \equiv x_{i,t-1}$ and enlarge the vector of endogenous variables to include $\tilde{x}_{i,t}$.

Aggregating across individuals, we have:

$$G(\theta)X_t = F(\theta)\int \mathcal{E}_{i,t}\left(\begin{pmatrix} x_{i,t+1} \\ X_{t+1} \end{pmatrix} | x_{i,t}, z_{i,t}, X_t, Z_t\right) d\Lambda(i) + L(\theta)Z_t \qquad (2.7)$$

$$Z_t = R(\theta)Z_{t-1} + \varepsilon_t \qquad (2.8)$$

We will refer to the economy with frictions as the triple $(H(\theta), \Lambda, \mathcal{E}_i)$ where

$$H(\theta) \equiv (vec(G(\theta)^T), vec(F(\theta))^T, vec(L(\theta))^T, vec(R(\theta))^T, vech(\Sigma_\varepsilon)^T).$$

We partition the vector θ into two subsets, (θ_1, θ_2) where θ_2 collects the parameters of the frictions. Thus, setting $\theta_2 = 0$, shuts down the frictions.

In an economy with no frictions, prices efficiently aggregate all the information. Thus there is no need to distinguish between individual and aggregate information when predicting aggregate state variables. When agents are rational, and the model is linear (or linearized) aggregate expectations for $(x_{i,t+1}^T, X_{t+1}^T)^T$ are as follows. For $w_{i,t} \equiv (x_{i,t}, X_{i,t}, z_{i,t}, Z_{i,t})$,

$$\begin{aligned}
\iint x_{i,t+1} p_i(x_{i,t+1}, X_{t+1}|w_{i,t}) d(x_{i,t+1}, X_{t+1}) d\Lambda(i) &= \\
\iint x_{i,t+1} p(x_{i,t+1}|w_{i,t}) d(x_{i,t+1}) d\Lambda(i) &= \\
\int P_{i,1} x_{i,t} d\Lambda(i) + P_2 X_t + P_3 \int P_{i,3} z_{i,t} d\Lambda(i) + P_4 Z_t & \quad \text{and} \\
\iint X_{t+1} p_i(x_{i,t+1}, X_{t+1}|w_{i,t}) d(x_{i,t+1}, X_{t+1}) d\Lambda(i) &= \\
\int X_{t+1} p(X_{t+1}|X_t, Z_t) d(X_{t+1}) &= \\
P_5 X_t + P_6 Z_t \equiv \mathbb{E}(X_{t+1}|X_t, Z_t) &
\end{aligned}$$

where $P_{j,j=1..6}$ are the coefficients of the linear projection. By Rational expectations and since the coefficients (G, F, L) are common across i, equilibrium consistency requires $P_{i,1} = P_1, P_{i,2} = P_2$ and therefore $P_1 + P_2 = P_5$ and $P_3 + P_4 = P_6$. Thus, as expected, aggregate conditional expectations collapse to $\mathbb{E}(X_{t+1}|X_t, Z_t)$.

The frictionless economy, $(H(\theta_1, 0), \Lambda, \mathbb{E})$, can be summarized by the equilibrium conditions:

$$G(\theta_1, 0) X_t = F(\theta_1, 0) \mathbb{E}_t(X_{t+1}|X_t, Z_t) + L(\theta_1, 0) Z_t \tag{2.9}$$

$$Z_t = R(\theta) Z_{t-1} + \varepsilon_t \tag{2.10}$$

The following assumption ensures the existence of a Rational Expectations equilibrium in a frictionless economy.

ASSUMPTION-EQ

There exist unique matrices $\underset{n_x \times n_x}{P^*(\theta_1, 0)}, \underset{n_x \times n_z}{Q^*(\theta_1, 0)}$ satisfying:

$$\begin{aligned}
(F(\theta_1, 0) P^*(\theta_1, 0) - G^*(\theta_1, 0)) P^*(\theta_1, 0) &= 0 \\
(R(\theta)^T \otimes F(\theta_1, 0) + I_z \otimes (-F(\theta_1, 0) P^*(\theta_1, 0) + G^*(\theta_1, 0))) vec(Q(\theta_1, 0)) &= -vec(L(\theta_1, 0))
\end{aligned}$$

such that $X_t = P^*(\theta_1, 0) X_{t-1} + Q^*(\theta_1, 0) Z_t$ is a competitive equilibrium.

These two conditions arise using the decision rule in the expectational system (2.9) and solving for the undetermined coefficients, see for example (Marimon and Scott 1998). Since the econometrician does not know the model with frictions (she knows the model up to $\theta_2 = 0$), we rearrange the equations of the economy with frictions into the known and the unknown part of the specification. Adding and subtracting the first order conditions of the frictionless economy, we have:

$$G(\theta_1,0)X_t = F(\theta_1,0)\mathbb{E}_t(X_{t+1}|X_t) + L(\theta_1,0)Z_t + \mu_t$$

where

$$\begin{aligned}
\mu_t \equiv & -(G(\theta) - G(\theta_1,0))X_t \\
& +(F(\theta) - F(\theta_1,0)) \int \mathcal{E}_{it}\left(\begin{pmatrix} x_{i,t+1} \\ X_{t+1} \end{pmatrix} |x_{i,t}, X_t\right) d\Lambda(i) \\
& +F(\theta_1,0)(\int \mathcal{E}_{it}\left(\begin{pmatrix} x_{i,t+1} \\ X_{t+1} \end{pmatrix} |x_{i,t}, X_t\right) d\Lambda(i) - \mathbb{E}_t(X_{t+1}|X_t)) \\
& +(L(\theta) - L(\theta_1,0))Z_t.
\end{aligned}$$

This system of equations cannot be solved without knowing the structure of μ_t. Nevertheless, we characterize the relationship between μ_t and a set of candidate decision rules that depend on the endogenous variables and some unobserved process, λ_t. The following proposition states sufficient conditions such that decision rules are consistent with μ_t.

Proposition 1. *Given:*

1. *The perturbed system of equilibrium conditions* (2.9) *where* $\mathbb{E}_t\mu_t \geq 0$
2. *A distorted aggregate decision rule* $X_t^* = X_t^{f,RE} + \lambda_t$ *where* $X_t^{f,RE} = \int x_{i,t}^{RE} d\Lambda(i)$ *is the Rational Expectations equilibrium of* $(H(\theta_1,0),\Lambda,\mathbb{E})$ *and*
3. *A* λ_t *vector process such that* $\lambda_t = \lambda_{t-1}\Gamma + \nu_t$ *for some real-valued* Γ*:*

If there exists a non-empty subset of Θ_1 *that satisfies*

$$\mathbb{E}_t(F(\theta_1,0)\Gamma - G(\theta_1,0))\lambda_t = -\mathbb{E}_t\mu_t$$

The following condition is satisfied for almost all subsets of $\sigma(Y_{t-1})$ *:*

$$\mathbb{E}_t(G(\theta_1,0)X_t^* - F(\theta_1,0)(\theta)X_{t+1}^* - L(\theta_1,0)Z_t) \geq 0 \tag{2.11}$$

Proof. See Appendix □

Proposition 4 states that, as long as there exists an admissible parameter vector $\theta_1 \in \Theta_1$ such that condition 3.6 holds, the decision rule $X_t^* = X_t^{f,RE} + \lambda_t$ generates the same restrictions as those implied using the perturbed (by μ_t) first order equilibrium conditions. Model incompleteness obtains because we only know the sign of λ_t. Moreover, we focus on parameter vectors that yield determinate and stable equilibria in the frictionless economy. This implies a restriction on the stochastic behavior of λ_t. Since there is limited information about λ_t, we are only able to identify sets of economic relationships out of average moment inequality conditions.

However, in certain situations, we may be interested in characterizing frictions over time (and not just on average), and therefore we need to obtain a conditional model that generates λ_t. This requires imposing more restrictions on the stochastic behavior of λ_t. We will show that restricting the distribution of aggregate shocks is enough. For every $\theta \in \Theta$, there is a multiplicity of corresponding conditional models which can be constructed with different distributions of shocks but give rise to the same $\mathbb{E}\lambda_t$.

To construct this family of models, we use the fact that the process λ_t causes a change in the measure implied by the frictionless model. As (Hansen 2013), we define a perturbation $\mathcal{M}_t$ such that for any measurable random variable W_t: $\mathbb{E}_t\mathcal{M}W_t = \mathbb{E}_t(W_t|M_f)$ (and vice versa, that is $\mathbb{E}_tW_t = \mathbb{E}_t(\tilde{\tilde{\mathcal{M}}}_tW_t|M_f)$). This representation is useful for two reasons. First, as we show in the proof of Proposition 4, interpreting the distortions as a change of measure provides a unified way of looking at frictions. Second, we can compute $\mathcal{M}_t$, for all t, and this can give us estimates of the wedges at any point of time.

We briefly explain how one can compute $\mathcal{M}_t$ - a full description is in the Appendix. Recall that in the linearized model, $\tilde{\lambda}_t$ measures the distance between the prediction of the frictionless model and the data, where the latter is assumed to be produced by a model with frictions. As shown in section 2, $\tilde{\lambda}_t$ is a function of the endogenous variables and the shocks. Without loss of generality assume that $\mathbb{E}\tilde{\lambda}_t > 0$. We look for a $\mathcal{M}_t$ that makes this expectation zero. By finding a $\mathcal{M}_t$ such that $\mathbb{E}\mathcal{M}_t\tilde{\lambda}_t = 0$ we are identifying $\mathcal{M}_t d\mathbb{P}(.)$, which is the density of the frictionless model that can be derived from the data by distorting the objective distribution, $\mathbb{P}$. Given $\mathcal{M}_t$, we can decompose $\mathbb{E}\tilde{\lambda}_t$ as: $\mathbb{E}\tilde{\lambda}_t \equiv \mathbb{E}\mathcal{M}_t\tilde{\lambda}_t + \mathbb{E}(1-\mathcal{M}_t)\tilde{\lambda}_t$. Therefore, to impose $\mathbb{E}\mathcal{M}_t\tilde{\lambda}_t = 0$ it suffices to set $\mathbb{E}(1-\mathcal{M}_t)\tilde{\lambda}_t = \mathbb{E}\tilde{\lambda}_t$. The term $1-\mathcal{M}_t$ determines the distortion at time t. To understand why using this decomposition is useful, notice that $\tilde{\lambda}_t$ is related to endogenous frictions but also to the unobservable shocks. As we show below, $1-\mathcal{M}_t$ is a time varying function of $\mathbb{E}\tilde{\lambda}_t$. Since the latter is an average, unobservable shocks are eliminated and $1-\mathcal{M}_t$ captures only the endogenous frictions.

In general, there are multiple $\mathcal{M}_t$ that satisfy the restriction $\mathbb{E}(1-\mathcal{M}_t)\tilde{\lambda}_t = \mathbb{E}\tilde{\lambda}_t$. This is exactly what we mean by a multiplicity of conditional models. In order to get a unique conditional model, we need to impose more restrictions on the stochastic behavior of $\mathcal{M}_t$ and therefore $\tilde{\lambda}_t$. To do this, we introduce a pseudo-distance metric $d(\mathcal{M}_t)$, which we minimize subject to the restriction $\mathbb{E}(\mathcal{M}_t-1)\tilde{\lambda}_t = \mathbb{E}\tilde{\lambda}_t$. The choice of the metric depends on the modelers' beliefs of the distribution of the shocks. Below we illustrate what happens when $d(\mathcal{M}_t) \equiv \frac{1}{2}(\mathcal{M}-1)^T(\mathcal{M}-1)$ which is consistent with the shocks having finite second moments [4]. Note that this distance metric is also used in the classic mean-variance frontiers in portfolio choice theory, or to compute Hansen-Jagganathan bounds. Intuitively, the minimization implies that we look for $\mathcal{M}_t$ that is consistent with our moment restrictions and has minimum variance [5]. Restricting the distribution of the shocks pins down a unique conditional model corresponding to a value of θ.

In general, $\mathcal{M}_t$ is a positive $\mathcal{F}_t-$ measurable random variable, unit expectation martingale, $\mathbb{E}(\mathcal{M}_{t+1}|\mathcal{F}_t) = \mathcal{M}_t, \mathbb{E}\mathcal{M}_t = 1$ [6]. We stack all $\mathcal{M}_t$ in a vector $\mathcal{M}$ and define $\tilde{\mathcal{M}} = \mathbf{1} - \mathcal{M}$ where $\mathbf{1}$ is the unit vector, which corresponds to the frictionless steady state of $\mathcal{M}_t$. The vector $\mathcal{M}$ satisfies the following program, where bold letters indicate vectors, $\tilde{\lambda}(Y_t;\theta)$ is the matrix containing the distortions for every t for every variable j, $\tilde{\lambda}_{j,t}$ and (π_1,π_2,π_3) are the corresponding Lagrange multiplier vectors.

$$\begin{aligned}
\max_{\mathcal{M}} \quad & -\frac{1}{2}\tilde{\mathcal{M}}^T\tilde{\mathcal{M}} \\
\text{subject to} \quad & \mathbf{1^T}\tilde{\mathcal{M}} = 0 \quad (\pi_1) \\
& \tilde{\mathcal{M}}^{\mathbf{T}}\tilde{\lambda}_j(Y_t;\theta) = \mathbf{1^T}\tilde{\lambda}_j(Y_t;\theta), \quad j = 1,..p \quad ,(\pi_{2,j}) \\
& \tilde{\mathcal{M}}^{\mathbf{T}}\tilde{\lambda}_j(Y_t;\theta) = \left[\mathbf{1^T}\tilde{\lambda}_j(Y_t;\theta)\right]_+, \quad j = p+1,..r \\
& \mathcal{M} \geq \mathbf{0} \quad (\pi_{3,t}, \forall t \in (1..T)
\end{aligned}$$

[4]For any random variable x and distorting density $\mathcal{M}_t$ by Cauchy Schwartz we have that $(\int x_t\mathcal{M}_t d\mathbb{P})^2 \leq \int x_t^2 d\mathbb{P}\int \mathcal{M}_t^2 d\mathbb{P}$. Minimizing the variance of the second term assumes that the variance of the first term exists.

[5]This is similar to the approach in generalized empirical likelihood settings in econometrics (i.e. (Newey and Smith 2004)).

[6]Being a martingale is a necessary condition for the distorted conditional expectation to be consistent with the Kolmogorov definition (Hansen and Sargent 2005).

The first constraint imposes unit expectation while the last constraint imposes non negativity of $\mathcal{M}$. The rest of the constraints impose the moment conditions. We present an analytical solution of this problem ignoring the third constraint. The set of complete Kuhn-Tucker conditions is in the Appendix. Letting $\bar{\lambda} \equiv T^{-1}\mathbf{1}^{\mathbf{T}}\tilde{\lambda}(Y;\theta)$, $\tilde{\tilde{\lambda}}(Y;\theta) \equiv \tilde{\lambda}(Y;\theta) - \bar{\lambda}(Y;\theta)$ and $\mathcal{V}(\tilde{\lambda}(Y;\theta)) \equiv T^{-1}\tilde{\tilde{\lambda}}(Y;\theta)^{\mathbf{T}}\tilde{\lambda}(Y;\theta)$ we have:

$$\tilde{\mathcal{M}} = \tilde{\tilde{\lambda}}(Y;\theta)\mathcal{V}(\lambda(Y;\theta))^{-1}\begin{bmatrix}\bar{\lambda}_1(Y;\theta)\\ [\bar{\lambda}_2(Y;\theta)]_+\end{bmatrix} = w_t\begin{bmatrix}0\\ [\bar{\lambda}_2(Y;\theta)]_+\end{bmatrix} \tag{2.12}$$

where $w_t \equiv \frac{\hat{Z}_t^2+\hat{\lambda}_t(\hat{\lambda}_t-\mathbb{E}\hat{\lambda}_t)}{\mathbb{V}(\hat{Z}_t)+\mathbb{V}(\hat{\lambda}_t))}$. The optimal $\mathcal{M}_t$ is a time varying function w_t of the average distortion over the sample, $\bar{\lambda}_2(Y;\theta)$. The weight w_t is a function of the relative variability of $\hat{\lambda}$, which is a function of the endogenous variables, and of $\hat{Z}_t$, which is function of the shocks Z_t.

We illustrate Proposition 4 when the econometrician does not know the capital adjustment cost function and the data have been generated using quadratic adjustment costs (as in Example 1). Note that here the distortion $\tilde{\mu}_t$ appears in the aggregate (linearized) Euler equation. We derive $\tilde{\lambda}_t$ for all endogenous variables.

Example 4. ***Analytical results for Example 1***

Given the first order conditions of the frictionless model, let $(\tilde{\lambda}_1,\tilde{\lambda}_2,\tilde{\lambda}_3,\tilde{\lambda}_4,\tilde{\lambda}_5,\tilde{\lambda}_6)$ *be the linearized distortions to* $(\tilde{K}_t,\tilde{K}_{t+1},\tilde{C}_t,\tilde{R}_t,\tilde{I}_t,\tilde{Y}_t)$. *Then, in this case* $\mathbb{E}_t(F(\theta_1,0)\Gamma - G(\theta_1,0))\tilde{\lambda}_t + \tilde{\mu}_t) = 0$ *is:*

$$\begin{bmatrix} 0 & -\beta\Gamma & 0 & -\beta r_{ss}\Gamma & 0 & 0\\ 1-\alpha & 0 & 0 & -1 & 0 & 0\\ 0 & 0 & -c_{ss} & 0 & -i_{ss} & y_{ss}\\ \delta-1 & 0 & 0 & 0 & 1 & 0\\ -\alpha & 0 & 0 & 0 & 0 & 1\\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix}\tilde{\lambda}_{1,t}\\ \tilde{\lambda}_{2,t}\\ \tilde{\lambda}_{3,t}\\ \tilde{\lambda}_{4,t}\\ \tilde{\lambda}_{5,t}\\ \tilde{\lambda}_{6,t}\end{bmatrix} = \begin{bmatrix}-\tilde{\mu}_{1,t}\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix}$$

where Γ *is the projection of* $\mu_{1,t+1}$ *on* $\mu_{1,t}$. *All of* $\tilde{\lambda}_{j,t}, j \in (1,2..6)$ *are determined by* $\tilde{\mu}_{1,t}$. *Letting*

$\Omega \equiv \Gamma\beta((\alpha-1)r_{ss}+1)$ *the solution is:*

$$\begin{bmatrix} \tilde{\lambda}_{1,t} \\ \tilde{\lambda}_{2,t} \\ \tilde{\lambda}_{3,t} \\ \tilde{\lambda}_{4,t} \\ \tilde{\lambda}_{5,t} \\ \tilde{\lambda}_{6,t} \end{bmatrix} = \frac{\tilde{\mu}_{1,t}}{\Omega} \begin{bmatrix} 1 \\ 1 \\ c_{ss}(-i_{ss}(1-\delta)+\alpha y_{ss}) \\ \alpha-1 \\ 1-\delta \\ \alpha \end{bmatrix}$$

Since the adjustment cost is unknown, one can only argue that $g(K_{t+1},K_t)$ *is increasing in* $I_r = \frac{K_{t+1}}{K_t}$*. In this case the non linear Euler equation becomes*

$$C_t^{-\omega} = \beta \mathbb{E} C_{t+1}^{-\omega}(1-\delta+R_{t+1}) - \beta \frac{\mathbb{E} C_{t+1}^{-\omega}}{K_t} g_{I_r,t}(I_{r,t} + \frac{C_{t+1}^{-\omega}}{\mathbb{E} C_{t+1}^{-\omega}})$$

where g_{I_r} *is the derivative of* $g(K_{t+1},K_t)$ *with respect to* I_r*. Therefore,* $\mu_{1,t} = -\beta \frac{\mathbb{E} C_{t+1}^{-\omega}}{K_t} g_{I_{r,t}}(I_{r,t} + \frac{C_t^{-\omega}}{\mathbb{E} C_{t+1}^{-\omega}}) < 0$*. Linearizing this equation does not change the sign of the conditional mean for the distortion. Moreover, we know that* $\lambda_{1,t} < 0$ *as the law of motion of capital is distorted by the adjustment cost. Given that we know sign*$(\mathbb{E}_t \tilde{\mu}_{1,t})$ *and sign*$(\mathbb{E}_t \tilde{\lambda}_{1,t})$*, we can derive sign*$(\mathbb{E}_t \tilde{\lambda}_{j,t})$ *for all* $j > 1$ *using equation (3.9).*

Moreover, given that we know the signs of the distortions to all of the variables, we can utilize a subset of these as moment restrictions. For example, we can use the moment inequalities for investment and capital:

$$\begin{aligned} \mathbb{E}\tilde{\lambda}_{5,t} &= \mathbb{E}(I_t - \tilde{I}_t^{\star})K_{t-1} \leq 0 \\ \mathbb{E}\tilde{\lambda}_{1,t} &= \mathbb{E}(K_t - \tilde{K}_t^{\star})K_{t-1} \leq 0 \end{aligned}$$

where $\tilde{I}^{\star}, \tilde{K}^{\star}$ *are the solution of the frictionless model. Correspondingly, we can estimate the perturbations* $\tilde{\mathcal{M}}_t$ *by substituting* λ_j *for* $\tilde{\lambda}_1, \tilde{\lambda}_5$ *in equation (3.7).*

For the rest of the paper we focus on the identification of the set of unconditional models, indexed by θ. In the next section we will show how qualitative survey data can be formally linked to the decision rules of the model, how they can be used to generate additional moment conditions and provide a characterization of the identified set of parameters in this case.

2.4 The Link to Qualitative Survey Data

To link survey data and DSGE model variables, one typically uses a measurement error approach, see (Del Negro and Schorfheide 2013), where the expected value of the survey variable is a proxy for a model variable. For example, inflation expectations taken from the Survey of Professional Forecasters have been linked to DSGE-model generated inflation expectations [7]. Let Y_t^o be the observed variable of interest and Y_t^m the model based conditional expectation of Y_t^o, $Y_t^m \equiv \mathbb{E}(Y_t^o|M,\mathcal{F}_t)$. Furthermore, let $\tilde{Y}_t$ be some cross sectional statistic in the survey e.g. the mean or the median. Then, the observation equation used is:

$$\tilde{Y}_t = \Lambda Y_t^m + u_t, \quad \mathbb{E}(u_t|Y_t^m) = 0$$

where Λ is either estimated or fixed. Combining with this observation equation with the restrictions implied by the model, the augmented set of restrictions is:

$$\begin{aligned} \mathbb{E}(Y_t^o - Y_t^m|\mathcal{F}_t) &= 0 \\ \mathbb{E}(\tilde{Y}_t - \Lambda Y_t^m|\mathcal{F}_t) &= 0 \end{aligned}$$

This paper introduces inequality restrictions on both equations of the form:

$$\mathbb{E}(Y_t^o - Y_t^m|\mathcal{F}_t) \geq 0 \tag{2.13}$$

$$\mathbb{P}(Y_t^o - Y_t^m \geq 0|\mathcal{F}_t) \geq \Xi_t \tag{2.14}$$

In the previous sections we have motivated equation 4.1. In this section we motivate equation 4.2. and how Ξ_t depends on survey data.

A general approach to the problem would be to treat equation 4.2 as an additional conditional moment restriction that can be constructed using Ξ_t:

$$\begin{aligned} \mathbb{E}(Y_t^o - Y_t^m|\mathcal{F}_t) &\geq 0 \\ \mathbb{E}\left((g(Y_t^o, Y_t^m, \Xi_t)|\mathcal{F}_t)\right) &\geq 0 \end{aligned}$$

The choice of function g depends on the type of additional data available. While our identifica-

[7] Similarly, in the single equation New Keynesian Phillips Curve, inflation expectations are substituted with survey data, see for example (Mavroeidis, Plagborg-Moller, and Stock 2014).

tion results hold for any function g and Ξ_t, we focus on survey data of categorical nature i.e. on representation 4.2.

Usually, qualitative survey data are in the form of aggregate statistics, where aggregation is performed over categories of answers to particular questions. We choose to represent answers as functions from the event space $\omega \in \Omega$ to the range of the random variables of interest. This is similar to the treatment in microeconometric studies, where it is assumed that the categorical variable is a weakly increasing function of a continuous latent variable. We divide the support of the control variables, which are assumed to be continuous, into a finite number of partitions which depends on the number of categories of answers available to the survey respondents. For example, if the question is of the type "How do you expect your financial situation to change over the next quarter" and the answer is trichotomous, i.e "Better", "Same" and "Worse", then the answers maps to partitioning end of period assets a_{t+2} into three sets: $a_{t+2} \in [a_{t+1} - \varepsilon, \infty)$, $a_{t+2} \in (a_{t+1} - \varepsilon, a_{t+1} + \varepsilon)$ and $a_{t+2} \in (-\infty, a_{t+1} - \varepsilon]$. Therefore, there is a measurable mapping from the categories of answers to the random variables relevant to the decision of each agent. Moreover, since the decision of each agent depends on the model she has in mind, there is a measurable mapping from answers to $\mathcal{E}_{i,t}(x_{t+1,i}|x_{i,t}, X_t)$. For this interpretation to hold, we need to assume that agents report their beliefs truthfully.

Denote by $\{S_{i,k,t}\}_{i \leq N}$ the survey sample over a period of length T, where i is the index over survey respondents, and k an index over $\mathcal{K}$, the class of questions that map to the same event on Ω. For example, if a respondent reports that she expects her "future savings" to deteriorate, she should also respond negatively to the question about her "future financial situation". Let C_l^k be the l^{th} categorical answer to question k and $\hat{\xi}_{i,t,k}$ the respondent's choice. Given some weights $\{w_l\}_{l \leq L}$ on each category, the available statistics are of the form:

$$\hat{\mathcal{B}}_t^k = \sum_{l \leq L} w_l \sum_{i \leq N} w_i 1(\hat{\xi}_{i,t} \in C_l^k)$$

The weight assigned on each individual response, w_i, corrects for discrepancies in representation between units, as it is done with the treatment of survey data by statistical agencies. For example, the response of a two hundred employee firm has to be adjusted so as to be comparable to a five employee firm. Without loss of generality, we assume a negative answer is equal to minus the positive one, and we therefore restrict our analysis to the case of $w_l \in \mathbb{R}_+$.

Since we have assumed truth telling, we can map the answer to the agent's beliefs i.e.

$h : \hat{\xi}_{i,t} \rightarrow \xi_{i,t} \equiv \mathcal{E}_{i,t}(x_{i,t+1}|x_{i,t}, X_t)$. Thus, the survey statistic is:

$$\hat{\mathcal{B}}_t^k = \sum_{l \leq L} w_l \sum_{i \leq N} w_i 1(\mathcal{E}_{i,t}(x_{t+1,i}|x_{i,t}, X_t) \in B_l) \tag{2.15}$$

where $B_l \equiv \{(x_{i,t}, X_t) \in \mathbb{R}^{2n_x} : h(x_{i,t}, X_t) \in C_l^k\}$. Since the conditional expectation is a function of the information set, B_l belongs to the partition of the support of individual $x_{i,t}$ (or aggregate X_t) that corresponds to category C_l^k. Representation 4.3 can then be linked to the model.

Recall that the perturbed solution of the model is:

$$\begin{aligned} \tilde{X}_t &= P^\star(\theta_1, 0)\tilde{X}_{t-1} + Q^\star(\theta_1, 0)\tilde{Z}_t + \tilde{\lambda}_t \\ &= P^\star(\theta_1, 0)\tilde{X}_{t-1} + Q^\star(\theta_1, 0)R(\theta)\tilde{Z}_{t-1} + Q^\star(\theta_1, 0)\varepsilon_t + \tilde{\lambda}_t \end{aligned}$$

Therefore, for every observable Y_t^o we have that:

$$Y_t^o \equiv Y_t^m + c'\varepsilon_t + \tilde{\lambda}_t \equiv Y_t^m + e_t + \tilde{\lambda}_t$$

If the distortion $\tilde{\lambda}_t$ is a positive function of the state variables, the macroeconomic event whose probability we want to characterize is $\mathcal{R}_t \equiv \{\tilde{\lambda}_t = Y_t^o - Y_t^m - e_t \geq 0\}$[8].

In section 3 we proposed a representation for $\tilde{\lambda}_t$ using the perturbation $\mathcal{M}_t$. This representation can be used when the user already has an estimator for the set Θ_I and is not appropriate for obtaining the identified set. Since $\tilde{\lambda}_t$ depends on the distance functional we minimize, the definition of $\mathcal{R}_t$ is non unique. Moreover, that representation of $\tilde{\lambda}_t$ requires information over the whole sample, while to construct $\mathcal{R}_t$ we need information at time t. This problem did not appear when the first set of restrictions was used (equation 4.1) because the identifying restrictions are unaffected by the representation of $\tilde{\lambda}_t$. In the second set of restrictions, the presence of the indicator function makes the representation of $\tilde{\lambda}_t$ important. We avoid this problem by treating $\tilde{\lambda}_t$ as an unobservable variable. To construct 4.2, we make an assumption on the probability of a linear combination of the vector of shocks being positive. This assumption is less stringent than those on the functional form of the distance functional in Section 3. Once $\hat{\Theta}_I$ is estimated, then $\tilde{\lambda}_t$ can be obtained.

[8]Strictly speaking, this characterization is true only for positively valued random variables. If the variables are transformed such that they take values on the whole real line, i.e. in deviation from steady state, then the proper event is $\mathcal{R}_t \equiv \{[Y_t^o - Y_t^m - e_t]X_{t-1} \geq 0\}$.

Since we deal with survey data, it is useful first to consider the event $\mathcal{R}_t$, at the individual level, $\mathcal{R}_{i,t} \equiv \{\lambda_{i,t} \geq 0\}$. In the proof of Proposition 6, we establish two facts. First, since the equilibrium conditions of the model with frictions depend on subjective conditional expectations, and the model with frictions is a smooth perturbation of the frictionless model, any probability statement on the subjective expectations translates to a probability statement on $\mu_{i,t}$. Second, any probability statement on $\mu_{i,t}$ is a probability statement on the solution of the model, and therefore on $\lambda_{i,t}$. Given representation (2.15), qualitative survey data have information on the quantiles of subjective conditional expectations of the agents. Therefore, survey data relate directly to the probability of observing a friction, $\mathbb{P}(\mathcal{R}_{i,t})$.

By treating shocks and frictions as unobservable, we can only construct $\mathbb{P}(\tilde{\mathcal{R}}_t) \equiv \mathbb{P}(Y_t^o - Y_t^m \geq 0)$. We show that $\mathbb{P}(\mathcal{R}_t)$ is bounded above by a constant times $\mathbb{P}(\tilde{\mathcal{R}}_t)$ and bounded below by $\mathbb{P}(\mathcal{R}_{i,t})$. This generates another moment inequality restriction. Proposition 6 shows that categorical survey data imply a lower bound on the probability of a distortion in an aggregate variable due to a friction. The bound depends on survey data and on the assumed distribution of the innovations to the shocks. The proposition deals with positive distortions; the same holds for negative distortions.

Proposition 2. *Let Λ be the distribution of agents in the economy and let $\hat{\mathcal{B}}_t^k$ be defined as in 4.3. If there exists an aggregate shock vector Z_t of length $p > 0$, with innovation ε_t such that $\mathbb{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\mathbb{V}(\varepsilon_t|\mathcal{F}_{t-1}) = \Sigma < \infty, \mathbb{P}-a.s$, then for a real valued $\underset{p\times 1}{c}$ such that $e_t = c'\varepsilon_t$ and $\kappa_t \equiv \mathbb{P}(e_t \geq 0|\mathcal{F}_{t-1})$, the following lower bound holds, $\mathbb{P}-a.s.$:*

$$\mathbb{E}(\mathbf{1}(Y_t^o - Y_t^m \geq 0) - \hat{\mathcal{B}}_t^k \kappa_t|\mathcal{F}_{t-1}) \geq 0 \tag{2.16}$$

Proof. See Appendix □

To obtain this result we assume that the shocks related to the friction are independent from the other structural shocks, and therefore independent from any linear combination of them ($c'\varepsilon_t$). For instance, in the capital irreversibility example, the marginal investment efficiency shock is assumed to be independent from the aggregate productivity shock. Moreover, it is typical to assume that the distribution of ε_t is symmetric around the origin, and therefore $\mathbb{P}(e_t \geq 0|\mathcal{F}_{t-1})$ is also easy to characterize. When this assumption is relaxed then characterizing $\mathbb{P}(e_t \geq 0|\mathcal{F}_{t-1})$ can be more difficult.

We also need assumptions about survey data and their econometric treatment. In order to cover cases which do not strictly belong to the 'indicator function' representation that categorical survey data imply, we consider a general representation which implies a choice of a statistic $\mathbb{S}$ over responses on question k, $\xi^k_{i,t}$. Examples of these statistics, for any function $\phi(.)$, can be averages, $\mathbb{S}_1(z) = (\mathbf{1}'\mathbf{1})^{-1}(\mathbf{1}'\phi(z))$ or other order statistics like $\mathbb{S}_2 = \phi(\min_i(z)), \mathbb{S}_3 = \phi(\max_i(z))$. In the case of categorical variables, $\phi(.) \equiv w_l w_i 1(.)$ and $\mathbb{S} \equiv \mathbb{S}_1 \circ \mathbb{S}_1$, therefore the statistic is $\Xi^k_t = \hat{\mathcal{B}}^k_t$. Our assumptions will be directly on Ξ^k_t:

ASSUMPTION-SD (Survey Data Assumptions)

1. $\{\Xi^k_t\}_{t\leq N}$ is a collection of random variables indexed by $t = 1..T$ with strong mixing coefficient α of size $-r/(r-1), r > 1$

2. $\mathbb{E}|\Xi^k_t|^{r+\delta} < \infty$

Lemma 1. ***Convergence of Survey Data Statistics***
Under Assumption SD and Theorem 2.3 in (White and Domowitz 1984), $|\frac{1}{T}\sum_t(\Xi^k_t - \mathbb{E}\Xi^k_t)| \overset{a.s}{\to} 0$

With regard to assumptions **SD-1** and **SD-2**, one can derive sufficient conditions for them to hold by characterizing the dependence of the state variables $(x_{i,t}, z_{i,t}, X_t, Z_t)$ across t and across i. This is because $\Xi_t \equiv \mathbb{S}(\xi_{i,t})$ and $\xi_{i,t}$ is a function of $x_{i,t}, z_{i,t}$, X_t, Z_t. We focus on parameter combinations that produce stable solutions, and on a linearized DSGE model that has a stationary VARMA(p,q) reduced form representation. Sufficient conditions for $\alpha-$ mixing of a VARMA(p,q) process would therefore be sufficient for our setup. Moreover, one needs to characterize how beliefs correlate across agents. For example, if $\mathbb{S} = \mathbb{S}_1$ and $r = 2$, a sufficient condition for assumption 2 to hold is that the absolute covariance of beliefs across agents is summable over i[9]. Milder conditions that allow for the sum to grow at a rate less than equal to N are also possible.

Apart from intrinsic randomness in survey data, there might be some additional noise in individual responses which can be attributed to classification error or heterogeneity. In the case of idiosyncratic noise, when $\mathbb{S} = \mathbb{S}_1$ the noise can be averaged out. If the noise is common across a subset of respondents, it is expected to generate bias in the identified bounds, especially if the additional survey based restrictions are non linear in Ξ_t. When the restrictions are linear in Ξ_t,

[9]Since $\xi^k_{i,t} \in (0,1)$, it has bounded moments. Moreover, $\mathbb{E}|\frac{1}{N}\sum_{i\leq N}\xi^k_{i,t}|^{2+\delta} \leq \mathbb{E}|\frac{1}{T}\sum_{i\leq N}\xi^k_{i,t}|^2$ and the latter is equal to $\frac{1}{N^2}\sum_{i\leq N}\mathbb{E}|\xi^k_{i,t}|^2 + \frac{1}{N^2}\sum_{j\neq i}\mathbb{E}|\xi^k_{i,t}\xi^k_{j,t}|$. The 1st component is bounded, as the second moment of $\xi^k_{i,t}$ is bounded for all i,t. For the second component, letting $h \equiv j-i$, $\mathbb{E}|\xi^k_{i,t}\xi^k_{j,t}| \equiv \alpha_{i,j}$ and $\alpha_{i,j} = \alpha_{j,i}$, $\frac{1}{N^2}\sum_{j\neq i}\mathbb{E}|\xi^k_{i,t}\xi^k_{j,t}| = \frac{2}{N}\sum_{h,h\geq 0}|\alpha_{i,i+h}|$. Summability of covariances of beliefs over different individuals at all times is sufficient for the 2nd component to converge to zero.

using as many questions as possible that relate to the same event might reduce this bias as long as the noise is independent across k. In the case of categorical survey data, $g(.,.)$ is separable in Ξ_t, and thus the bias can be reduced.

In the next section we analyze the implications of the restrictions of Proposition 6 for model identification.

2.5 Identification And Estimation

In the previous sections we have motivated how inequality restrictions arise from robust theoretical predictions and the relation between agents subjective expectations and qualitative survey data. In this section we provide a formal treatment of the identification of a linear(ized) Dynamic Stochastic General Equilibrium (DSGE) model based on these restrictions. First, we illustrate how our statistical representation of a DSGE model relates to the state space representation that is typically used for estimation. Building on (Komunjer and Ng 2011), we will show necessary and sufficient conditions for partial identification of the model arising from the theoretical moment inequalities. We also show conditions under which the survey data inequalities provide additional information and therefore a sharper identified set.

We base the analysis on the innovation representation of the solution to the linear(ized) DSGE model. This is the natural representation to use when there are differences in information between economic agents and the econometrician, as it takes into account that not all the state variables relevant the decision of agents are observable. We consider the following class of models:

$$
\begin{aligned}
\hat{X}_{t+1|t} &= \underset{n_X \times n_X}{A(\theta)} \hat{X}_{t|t-1} + \underset{n_X \times n_X}{K_t(\theta)} a_t \\
Y_t^o &= \underset{n_Y \times n_x}{C(\theta)} \hat{X}_{t|t-1} + a_t
\end{aligned}
$$

where $K_t(\theta)$ is the Kalman gain and a_t is the one-step ahead forecast error which could be derived from the state space representation:

$$
\begin{aligned}
X_{t+1} &= \underset{n_X \times n_X}{A(\theta)} X_t + \underset{n_X \times n_X}{B(\theta)} \varepsilon_{t+1} \\
Y_t &= \underset{n_Y \times n_X}{C(\theta)} X_t + \lambda_t = \underset{n_Y \times n_X n_X \times n_X}{C(\theta) A(\theta)} X_{t-1} + \underset{n_Y \times n_X n_X \times n_X}{C(\theta) B(\theta)} \varepsilon_t + \lambda_t \\
&= \underset{n_Y \times n_X}{\tilde{C}(\theta)} X_{t-1} + \underset{n_Y \times n_X}{D(\theta)} \varepsilon_t + \lambda_t
\end{aligned}
$$

where ε_t is the innovation to the shock vector Z_t. By construction,

$$a_t = \lambda_t + C(\theta)A(\theta)(X_{t-1} - \hat{X}_{t-1|t-1}) + C(\theta)B(\theta)\varepsilon_t$$

Therefore, the forecast error is a combination of the true innovations to the information sets of the agents, ε_t, the estimation error of the state variable, $X_{t-1} - \hat{X}_{t-1,t-1}$, and the frictions, λ_t. Let $N(\theta) \equiv vec(A(\theta)', B(\theta)', C(\theta)')$, and assume that $\mathbb{E}(\varepsilon_t|\sigma(\mathcal{F}_t)) = 0$, and $\mathbb{E}(\varepsilon_t\varepsilon_s'|\sigma(\mathcal{F}_t)) = \mathbf{1}(s=t)\Sigma_{\varepsilon_t}$, where $\Sigma_{\varepsilon_t} \succ 0$.

Given $\mathbb{E}(\lambda_t|\sigma(\mathcal{F}_{t-1})) \geq 0$, we define the following conditional moment restriction:

$$\mathbb{E}(Y_t^o - \underset{n_Y \times n_x}{C(\theta)}\hat{X}_{t|t-1}|\sigma(\mathcal{F}_{t-1})) \geq 0$$

For any function $\phi(.)$ of a random vector $\mathbf{Y}_{t-1}$ that belongs to the information set of the econometrician, the following holds:

$$\mathbb{E}(Y_t^o - \underset{n_Y \times n_x}{C(\theta)}\hat{X}_{t|t-1})\phi(\mathbf{Y}_{t-1}) = \mathbb{E}\mathcal{V}(\mathbf{Y}_{t-1})\phi(\mathbf{Y}_{t-1}) \geq 0$$

for a random function $\mathcal{V}(\mathbf{Y}_{t-1}) \in [0,\infty]$. In order to study the properties of these estimating equations we need to make assumptions about the local identification of Θ_0, given the value of $\mathbb{E}\mathcal{V}(\mathbf{Y}_{t-1})\phi(\mathbf{Y}_{t-1})$. We resort to sufficient conditions that make the mapping from θ to the solution of the model regular, and thus assume away population identification problems (see, for example (Canova and Sala 2009)). We assume that Θ belongs to a compact subset of $\mathbb{R}^{n_\theta}$. Since certain parameters are naturally restricted, e.g. discount factors, persistence parameters or fractions of the population, and others cannot take excessively high or low values, assuming compactness is innocuous. We also need to acknowledge that due to *cross - equation restrictions*, which we denote by $\mathfrak{L}(\theta) = 0$, the number of observables used in the estimation need not be equal to the cardinality of Θ, i.e., $n_y < n_\Theta$. (Komunjer and Ng 2011) provide the necessary and sufficient conditions for local identification of the DSGE model from the auto-covariances of the data. We reproduce them below, with the minor modification that Assumption **LCI-6** holds for any element of the identified set Θ_0.

ASSUMPTION -LCI (Local Conditional Identification)

1. Θ is compact and connected

2. (Stability) For any $\theta \in \Theta$ and for any $z \in \mathbb{C}$, $det(zI_{n_X} - A(\theta)) = 0$, implies $|z| < 1$

3. For any $\theta \in \Theta$, $D(\theta)\Sigma_e D(\theta)'$ is non-singular

4. For any $\theta \in \Theta$, (i) The matrix $(K(\theta)A(\theta)K(\theta)..,A(\theta)^{nx-1}K(\theta))$ has full row rank and $(C(\theta)'A(\theta)'C(\theta)'..,A(\theta)'^{nx-1}C(\theta)')'$ has full column rank.

5. For any $\theta \in \Theta$, the mapping $N : \theta \mapsto N(\theta)$ is continuously differentiable

6. Rank of matrix $\Delta^{NS}(\theta)$ is constant in a neighborhood $\theta_0 \in \Theta_I$ and is equal to $n_\theta + n_x^2$ [10]

Lemma 2. *Given $\mathcal{V}_i(.) \in [\underline{\mathcal{V}}(.), \bar{\mathcal{V}}(.)]$, and Assumption **LCI**, θ is locally conditionally identified at a θ_0 in Θ_I from the auto-covariances of Y_t. Consequently, θ is locally conditionally identified at any θ_0 in Θ_I.*

Proof. See Appendix □

In order to characterize the identified set, we also need to make assumptions on the correct specification of the moment conditions we use. Correct specification is important as it implies a non-empty identified set that covers the true θ_0. We give two characterizations of this concept, one that involves survey data and the other macroeconomic aggregates. Both have to be consistent with the model's predictions.

Definition 1. ***Correct Specification***
CS-I: *Let $\mathcal{K}$ be the class of questions that relate to the same event. For any question $K \in \mathcal{K}$, there does not exist a $K' \in \mathcal{K}$ such that $\bigcap_{k=\{K,K'\}} \left(\bigcup_{l \in L} B_{l,k}\right) = \emptyset, \mathbb{P} - a.s$*
CS-II: *For $\mathcal{R}_t \equiv \{Y_t^o - Y_t^m - e_t \geq 0\}$, $\exists \theta \in \Theta : \mathbb{P}(\mathcal{R}_t|\sigma(\mathcal{F}_t)) > 0$.*

The first condition, **CS-I**, implies that conditional on information at time t, the events identified by different survey questions cannot be mutually exclusive[11]. The second condition **CS-II** implies that for the model to be correctly specified, there must exist a subset of Θ such that

---

[10]

$$\Delta^{NS}(\theta_0) = \left(\frac{\partial \delta^{NS}(\theta, I_{n_x})}{\partial \theta}, \frac{\partial \delta^{NS}(\theta, I_{n_x})}{\partial vecT}\right)|_{\theta=\theta_0}$$

where

$$\delta^{NS}(\theta, T) = (vec(TA(\theta)T^{-1})^T, vec(TK(\theta))^T, vec(C(\theta)T^{-1})^T, vech(\Sigma_\alpha(\theta))^T)^T$$

[11] In the Appendix we plot time series of average responses to questions that covers approximately similar events. Mutual consistency should reveal as strong co-movement in the series.

the event we are interested to identify, $\mathcal{R}_t$, has positive conditional (and therefore unconditional) probability. In Lemma 11 in the Appendix we show that the two conditions, are equivalent. If either does not hold, the identified set is empty. Therefore, under correct specification, the identified set is non empty, $\Theta_I \cap \Theta \neq \emptyset$.

All of the identification results assume the existence of appropriate instruments to construct unconditional moment restrictions from the conditional moment inequalities [12]. Such instruments can be either past data or past state variables constructed with the Kalman filter. By construction, the latter are uncorrelated with current information, but they might be noisy.

With regard to the survey data moment conditions, we need to show their validity as "supernumerary", see (Bontemps, Magnac, and Maurin 2012), i.e. that they are useful in potentially reducing the identified set by making the set of admissible $\mathcal{V}(\mathbf{Y}_{t-1})$ smaller. We cannot strictly follow the approach of (Bontemps, Magnac, and Maurin 2012) here for two reasons. First, their characterization of the identified set is done through the support function of Θ_I[13], which requires the identified set to be convex. Since the stability conditions and the cross equation restrictions introduce nonlinearities on $\mathfrak{L}(\theta)$ and therefore nonlinear restrictions on Θ, the identified set is not necessarily convex. Second, we deal with additional moment conditions coming from survey data and not additional instruments. The main argument however, remains the same: if the additional moment conditions carry information, then the identified set will be weakly smaller.

Recall that the number of moment conditions we use for estimation depend on the number of observables. Assumption **LCI-6** requires that there has to be enough (or the right kind) of observables such that a rank condition is satisfied. In our case, the number of observables used determines the number of first order conditions of the minimization. The minimum number (r) of observables required such that conditional identification is achieved (Lemma 8 is satisfied) maps to the necessary first order conditions. For example, if we have Y_1 and Y_2 to estimate the model, and we only need Y_1 to conditionally identify θ, then the $n_\theta \times 1$ first order conditions arising from Y_1 will be the necessary conditions. The rest of the conditions, those arising from Y_2 and from survey data are then supernumerary.

Let $m_{\alpha,t}(\theta)$ and $m_{\beta,t}(\theta)$ denote the necessary and supernumerary moment functions for identifying Θ, where $\mathbb{E}(m_{\alpha,t}(\theta)|\mathbf{Y}_{t-1}) = \mathcal{V}_\alpha(\mathbf{Y}_{t-1}) \in [\underline{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1}), \bar{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1})]$ and

[12] Since we have conditional moment inequalities, we can in principle construct an infinite number of moment inequality restrictions, so using a finite number of them involves information loss, and therefore the identified set is not as sharp as possible. Recent work in the literature proposes constructing instrument functions to avoid this information loss, see for example (Andrews and Shi 2013). We nevertheless do not pursue this in this paper.

[13] The support function, which is $\sup(q^T\Theta), \forall q \in \mathbb{R}^{n_\theta}$, can fully characterize any convex set Θ.

$\mathbb{E}(m_{\beta,t}(\theta)|\mathbf{Y}_{t-1}) = \mathcal{V}_\beta(\mathbf{Y}_{t-1}) \in [\underline{\mathcal{V}}_\beta(\mathbf{Y}_{t-1}), \bar{\mathcal{V}}_\beta(\mathbf{Y}_{t-1})]$. Comparing these general bounds to the ones implied by the DSGE and survey data restrictions, $\underline{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1}) = \underline{\mathcal{V}}_\beta(\mathbf{Y}_{t-1}) = 0$ and $\bar{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1}) = \bar{\mathcal{V}}_\beta(\mathbf{Y}_{t-1}) = \infty$. Due to the boundedness of Θ and the cross-equation and stability restrictions, the actual lower and upper bounds are likely to lie strictly within $[0,\infty]$ for every moment condition. Let $\phi(.)$ be any $\mathbf{Y_{t-1}}-$ measurable function for which $\hat{m}_{\alpha,t}(\theta) := m_{\alpha,t}(\theta)\phi(\mathbf{Y_{t-1}})$ and $\hat{m}_{\beta,t}(\theta) := m_{\beta,t}(\theta)\phi(\mathbf{Y_{t-1}})$, $\hat{\mathbf{m}}_\alpha(\theta)$ and $\hat{\mathbf{m}}_\beta(\theta)$ the corresponding vectors, and $\bar{\mathbf{m}}_\alpha(\theta)$ and $\bar{\mathbf{m}}_\beta(\theta)$ the vector means. Denote by Q_α the projection matrix for projecting on $W_\alpha^{\frac{1}{2}T}\hat{\mathbf{m}}_\alpha(\theta)$, where W is a real valued, possibly random weighting matrix, diagonal in (W_α, W_β).

The following proposition shows that as long as there is a proportion of agents who face frictions, then survey data sharpen the identified set.

Proposition 3. *Characterization of the identified set:*

Given

1. *A sample of survey data* $\{\hat{\mathcal{B}}_t\}_{t\leq N}$
2. ***CS-I***
3. $U_t \in [\underline{U}, \bar{U}]$ *where*

$$\underline{U} \equiv (W_\alpha^{\frac{1}{2}T}\underline{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1}) + W_\beta^{\frac{1}{2}T}\underline{\mathcal{V}}_\beta(\mathbf{Y}_{t-1}))\phi(\mathbf{Y}_{t-1}) \quad \textit{and}$$
$$\bar{U} \equiv (W_\alpha^{\frac{1}{2}T}\bar{\mathcal{V}}_\alpha(\mathbf{Y}_{t-1}) + W_\beta^{\frac{1}{2}T}\bar{\mathcal{V}}_\beta(\mathbf{Y}_{t-1}))\phi(\mathbf{Y}_{t-1})$$

If $\hat{\mathcal{B}}_{k,t} \xrightarrow{p} B > 0$, *then the following condition holds non trivially*

$$\mathbb{E}Q_\alpha^\perp\left(W_\beta^{\frac{1}{2}T}\hat{m}_{\beta,t}(\theta) - \hat{U}_t\right) = 0 \tag{2.17}$$

Consequently, $\Theta_I' \subset \Theta_I$.

Proof. See the Appendix □

The main argument behind Proposition 9 is the following. Suppose that the necessary moment conditions have no common information with the supernumerary conditions and that $W = I_{n_\alpha+n_\beta}$. From the minimization of $\mathbb{E}\frac{1}{2}(\bar{\mathbf{m}} - \bar{U}_t)^T(\bar{\mathbf{m}}(\theta) - \bar{U}_t)$ where $\bar{\mathbf{m}}(\theta) \equiv (\bar{\mathbf{m}}_\alpha(\theta), \bar{\mathbf{m}}_\beta(\theta))^T$, the first order condition is

$$\mathbb{E}(\bar{\mathbf{m}}_\alpha(\theta) + \bar{\mathbf{m}}_\beta(\theta) - \bar{U}_t) = 0$$

which can be rewritten as:

$$\mathbb{E}((\bar{\mathbf{m}}_{\alpha}(\theta) - \overline{Q_{\alpha}U_t}) + (\bar{\mathbf{m}}_{\beta}(\theta) - \overline{Q_{\alpha}^{\perp}U_t})) = 0$$

By construction the two parts of the left hand side of the expression are independent, and therefore both have to be zero.

$$\mathbb{E}(\bar{\mathbf{m}}_{\alpha}(\theta) - \overline{Q_{\alpha}U_t}) = 0 \tag{2.18}$$

$$\mathbb{E}(\bar{\mathbf{m}}_{\beta}(\theta) - \overline{Q_{\alpha}^{\perp}U_t}) = 0 \tag{2.19}$$

Notice that, by construction, the set of necessary moment conditions in 5.3. must have full rank, and this establishes a one-to-one mapping from Θ to the domain of variation of U_t, $[\underline{\mathrm{U}}(\mathbf{Y}_{t-1}), \bar{U}(\mathbf{Y}_{t-1})]$. Thus, there exists an inverse operator $\mathcal{G}_{\alpha}$ such that $\theta = \mathcal{G}_{\alpha}(U_t, \mathbb{P})$. Plugging this expression for θ in 5.4, we get that

$$\mathbb{E}\bar{\mathbf{m}}_{\beta}(\mathcal{G}_{\alpha}(U_t, \mathbb{P})) = \mathbb{E}\overline{Q_{\alpha}^{\perp}U_t}$$

This is a restriction on the values that U_t can take in addition to the ones implied by the necessary conditions. A restriction on U_t implies a restriction on the admissible Θ_I given the one-to-one relationship in 5.3. Notice that when the supernumerary conditions do not add any additional information, i.e. $m_{\alpha,t}(\theta) \equiv m_{\beta,t}(\theta)$, the restriction collapses to $Q_{\alpha} = Q_{\alpha}^{\perp} = \frac{1}{2}$.

We illustrate below how survey data constraints provide information in the context of Example 1.

Example 5. ***Identification in the case of capital adjustment costs***

We assume that the representative firm faces adjustment costs with iid probability B_t which is a random variable with mean B. As long as B is positive, the conditional mean of $\tilde{\lambda}_t$ is the same as the one derived in example 1. We focus on identification using the aggregate capital accumulation equation. Denote the solution of the frictionless model as: $K_t = \varphi_k(\theta)K_{t-1} + \varphi_z(\theta)Z_t$. Therefore, $\mathbb{E}_{t-1}K_t = \varphi_k(\theta)K_{t-1}$. Let ζ_t denote an instrument and (ζ, K, K_{-1}, B) the vectors containing data on $(\zeta_t, K_t, K_{t-1}, B_t)$. The two identifying conditions are:

$$\mathbb{E}\zeta^T(K - \varphi_k(\theta)K_{-1}) = v_1, v_1 \leq 0 \tag{2.20}$$

$$\mathbb{E}\zeta^T(\mathbb{1}(K - \varphi_k(\theta)K_{-1} \leq 0) - B) = v_2, v_2 \geq 0 \tag{2.21}$$

Rearranging the first equation and letting $\varphi_k \equiv \mathbb{E}(\zeta^T K_{-1})^{-1}\mathbb{E}\zeta^T K$, *we get a lower bound for* $\varphi_k(\theta)$*:*

$$\varphi_k(\theta) = \varphi_k - \mathbb{E}(\zeta^T K_{-1})^{-1} v_1 \geq \varphi_k$$

Similarly, from the second equation, we have:

$$\mathbb{E}(\zeta^T\zeta)^{-1}\mathbb{E}\zeta^T \mathbb{1}(K - \varphi_k(\theta)K_{-1} \leq 0) = \mathbb{E}(\zeta^T\zeta)^{-1}\mathbb{E}\zeta^T B + \mathbb{E}(\zeta^T\zeta)^{-1} v_2$$

Letting $\phi_s(\theta) \equiv \mathbb{E}(\zeta^T\zeta)^{-1}\mathbb{E}\zeta^T \mathbb{1}(K - \varphi_k(\theta)K_{-1} \leq 0)$ *and* $\varphi_s \equiv \mathbb{E}(\zeta^T\zeta)^{-1}\mathbb{E}\zeta^T B$, *we have:*

$$\phi_s(\theta) = \varphi_s + \mathbb{E}(\zeta^T\zeta)^{-1} v_2 \geq \varphi_s \tag{2.22}$$

Using the first equation only, the interval we are able to identify is $\varphi_k(\theta) \in (\varphi_k, \bar{\varphi})$ *where* $\bar{\varphi}$ *is the natural upper bound of* Φ*. Let* $\hat{\mathbf{m}}_\beta$ *be the vector of observations of* $m_{\beta,t} := \zeta_t \mathbb{1}(K_t - \varphi_k(\theta)K_{t-1} \leq 0)$ *and* $\hat{\mathbf{m}}_\alpha$ *the vector of observations of* $m_{\alpha,t} := \zeta_t(K_t - \varphi_k(\theta)K_{t-1})$*. Let* $\hat{\gamma}$ *denote the estimated coefficient of the projection of* $\hat{\mathbf{m}}_\beta$ *on* $\hat{\mathbf{m}}_\alpha$ *and* γ *the population coefficient. Then* $m_{\beta,t}$ *is:* $\hat{m}_{\beta,t} = \gamma\hat{m}_{\alpha,t} + u_t$ *where* $\mathbb{E}(u_t|m_{\alpha,t}) = 0$*. Taking the unconditional expectation, we have* $v_1 = \gamma v_2$*, and therefore the following relation between model based and data based quantities:*

$$\varphi_s(\theta) - \varphi_s = |\gamma|(\varphi_k(\theta) - \varphi_k) \geq 0 \tag{2.23}$$

Equation (2.23) *shows how the bounds on* θ *present in the capital accumulation equation can be refined by considering additional survey information. If* $\gamma = 0$*, the additional bound in* (2.22) *is as informative as possible. When* $\gamma > 0$*, some of the additional information is already present in the original condition, making the refinement on the bound less pronounced.*

It is important to stress that adding additional data will not give point identification as long as there is heterogeneity and unobservable shocks. The convolution of two unobservables, the aggregate distortions and shocks, in fact does not allow for point identification.

Corollary 2.1. ***Impossibility of Point Identification****: If* $\dim Z_t > 0$ *or* $\Lambda(i,t)$ *is non degenerate, then* Θ_I *is not a singleton*

Proof. See Appendix □

The fact that qualitative survey data can potentially reduce the class of admissible theoretical models in a data driven way, makes them useful to validate models with parametric specification of frictions. However, care is required as assumptions need to be satisfied in order to make the most of the information contained in qualitative survey data.

2.6 Testing Parametric Models of Frictions

We showed in the previous sections how to obtain the set of parameters Θ_I that is consistent with frictions. Given the definition of $\lambda(Y_t, \theta)$ in Section 3, a plug-in set estimate of the average distortion (wedge) in a macroeconomic variable Y_t is $\mathbb{E}\lambda(Y_t, \theta_I(Y))$. This estimate can be used to validate a particular parametric model, since a DSGE model with a particular specification of frictions can produce wedges that fail to lie in the identified set. If survey data are informative, the identified set is sharper, that is, for $\Upsilon \equiv (Y_t, \Xi_t)$, $\theta_I(Y) \subset \theta_I(\Upsilon)$. Therefore there are values $\theta \in \Theta$ such that theory based restrictions are satisfied, but are not admissible, in the sense that they are not consistent with survey data. Based on this idea, we propose a test statistic and show its asymptotic size and power against fixed alternatives. We motivate its use through the example on capital adjustment costs.

Example 6. ***Validating capital adjustment costs***

Capital adjustment costs can be theoretically motivated using different setups see, for example, (Wang and Wen 2012). All setups can be equivalently represented with an idiosyncratic shock to the marginal efficiency of investment. Having such a shock implies time variation in Tobin's q, Q_t, which is the Lagrange multiplier on the capital accumulation equation. Rewrite the capital accumulation equation as $K_{t+1} = (1-\delta)K_t + \psi(\hat{\imath}_t)K_t$ where $\hat{\imath}_t \equiv \frac{I_t}{K_t}$ and $\psi(\hat{\imath}_t)$ is a function of the efficiency shock. The first order conditions relevant for investment and capital in the example are the following:

$$
\begin{aligned}
Q_t &= 1/\psi'(\hat{\imath}_t) \\
Q_t &= \beta \mathbb{E}_t \frac{C_{t+1}^{-\omega}}{C_t^{-\omega}} (R_{t+1} + (1-\delta)Q_{t+1} + Q_{t+1}\psi(\hat{\imath}_{t+1}) - i_{t+1}) \\
K_t &= (1-\delta)K_{t-1} + \psi(\hat{\imath}_t)K_t
\end{aligned}
$$

When $\psi(\hat{\imath}_t) = \hat{\imath}_t$ the conditions are the same as in the frictionless economy. In example 1, the economy has additive adjustment costs, $K_{t+1} = (1-\delta)K_t + I_t - \frac{\phi}{2}(\frac{K_{t+1}}{K_t} - 1)^2 K_t$ which amounts to choosing $\psi(\hat{\imath}_t) = \hat{\imath}_t - \frac{\phi}{2}(\frac{K_{t+1}}{K_t} - 1)^2$.

Suppose one erroneously assumes that capital adjustment costs are due to financing frictions. We adopt one of the specifications of (Wang and Wen 2012), where there are heteroge-

neous firms which differ only in the marginal investment efficiency. Efficiency follows a Pareto distribution across firms, with shape parameter η. *Investment must be financed before production takes place, and thus firms must issue a one period loan. The lender then lends a fraction* θ *of the value of capital, which on aggregate, amounts to a restriction on aggregate investment of the following form,* $I_t = \theta Q_t^{\eta+1} K_{t-1}$. *This results in* $\psi(\hat{\iota}_t) = \frac{\eta}{\eta-1}\theta^{\frac{1}{\eta+1}}\hat{\iota}_t^{\frac{\eta}{\eta+1}}$. *Since the reduced form of the model is the same as the reduced form of the original economy, despite the fact that the functional form for* $\psi(.)$ *is different, the two models are observationally equivalent.*

We simulate data from the true economy when there is a 0.8 unconditional probability for the representative firm to face an adjustment cost and we generate survey data by adding an iid uniformly distributed shock u to generate a time varying conditional probability. Let the productivity shock be Normally distributed.

We conduct the following experiment. We first estimate the identified set of frictions using only the robust inequality restrictions on investment and capital derived in example 5. Then re-estimate the identified set by adding the survey data restrictions. In particular, for this simple example, the additional moment inequality is: $\mathbb{E}(1(K_t^o - \tilde{K}_t) - \frac{1}{2}(0.8 + u_t)) \geq 0$. *Finally, we estimate the misspecified model, obtain a point estimate, and produce estimates of frictions by plugging the set and point estimates in* $\mathbb{E}\lambda(Y_t, \theta_I(Y))$.

Figures $1.1 - 1.2$ *plot posterior draws of the estimated wedges for capital using the misspecified model (blue) and the robust set estimates (red) without and with survey data. We also report the point corresponding to the true model (dashed line) and the point estimate corresponding to the mis-specified model (green circle). Clearly, the true estimates lies within the 'red' set, while the 'blue' estimates are far away, indicating misspecification. In Figure 2.2, the use of survey data is to make inference sharper by further constraining the admissible frictions. The proposed model of input financing frictions, although having the same reduced form as the true model, it is 'too far' from it, in the sense that stability restrictions on the parameters result in a point estimate that is far from the red set, and we can therefore reject the model by only using the economic theory based moment inequalities. It can be the case though that the economic moment conditions are not informative enough to decide whether the complete model with frictions is valid. This depends on how much misspecification there is, and how much sampling uncertainty. By using survey data, we obtain a sharper identified region, leading to a more precise inference about whether the proposed mechanism is valid.*

The Wald statistic we use tests whether the expected distance from the point estimate of

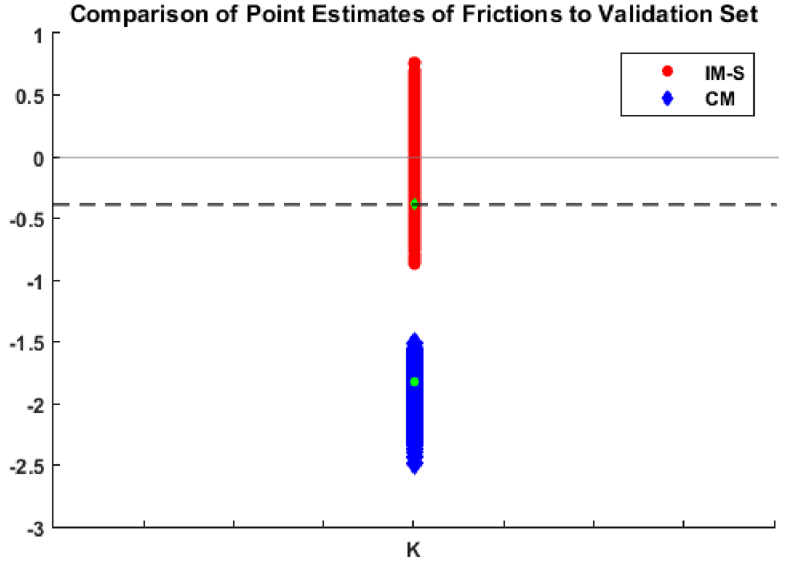


Figure 2.1: Pseudo - Posterior Capital Wedge estimates without Survey data

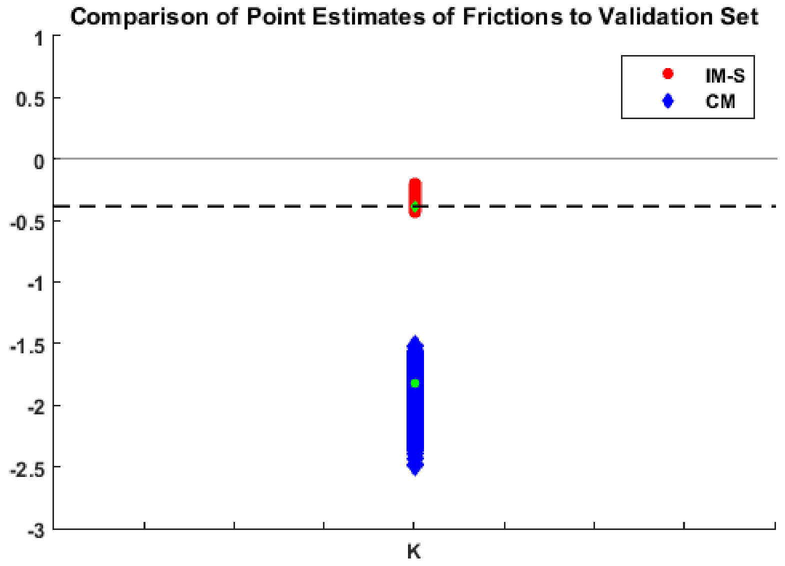


Figure 2.2: Pseudo - Posterior Capital Wedge estimates with Survey data

the parametric model (green circle) to the (red) identified set is different than zero for all (or some of) the observables. The statistic is:

$$\mathcal{W}_t = \left(\sqrt{t} \inf_{\lambda_s \in \lambda(\hat{\Theta}_s)} ||\mathcal{V}^{-\frac{1}{2}}(\lambda_s - \lambda_p^\star)|| \right)^2$$

where λ is the estimated friction obtained using either the identified point in the parametric model case, λ_p, or the identified set in the robust case, λ_s. Individual frictions are weighted by their respective estimate of standard deviation. The statistic measures the Euclidean distance between the wedge that arises in the parametric model, and the set of admissible wedges, adjusted for estimation uncertainty. The Null hypothesis we seek to test is that $\lambda(\theta_p) \in \lambda(\Theta_s)$; the alternative is $\lambda(\theta_p) \notin \lambda(\Theta_s)$. Under the Null , $\mathbb{E}\mathcal{W}_t = 0$; under the alternative, $\mathbb{E}\mathcal{W}_t > 0$.

To get critical values we use the following bootstrap procedure, where superscript '$\star$' denotes the bootstrapped data:

1. Obtain B samples of length T of data, $\mathbf{Y}^\star_{t,l}$, for $t \leq T, b \leq B$ using a preferred consistent block sampling scheme with block length l, e.g. (Politis and Romano 1994).

2. For each sample, obtain $\hat{\theta}_p$ and $\hat{\Theta}_s$, and compute $\min_{f_s} \sum_{j \in 1..k} \tilde{\mathcal{V}}_j^{-1}(\lambda_{p,j}(\mathbf{Y}^\star) - \lambda_{s,j}(\mathbf{Y}^\star))^2$ where j is the index of observables included in the test

3. Compute the $1-\alpha$ quantile of the empirical distribution of $\hat{\mathcal{W}}_b, b \leq B$

This bootstrap test, under certain regularity conditions, is consistent and has asymptotic power equal to one against fixed alternatives. The regularity conditions we need are the following:

ASSUMPTION -R (Regularity conditions)

Let D and q be $Y-$ measurable functions, continuous in θ w.p.1 and $\tilde{q}(.;\theta) \equiv \tilde{q}(.;\theta) - \bar{q}(.;\theta)$ such that:

1. For any $\mathbf{Y}$, and any $\theta \in \Theta$, $\sqrt{T}\tilde{q}(\mathbf{Y};\theta) \rightarrow_d \mathcal{N}(0,\Omega)$

2. $\sup_{\theta \in \Theta} D_n(\theta) \rightarrow_d D$ where D is positive definite

3. The above statements hold either for the original sample, or for any sequence of bootstrap samples, conditional on almost all original sample paths.

Proposition 4. ***Size and Power of Bootstrap Test for Fixed Alternative***

Given:

1. *Assumptions **SD** and **R***

2. $(l,T,B) \rightarrow \infty$ *and* $\frac{l}{T} \rightarrow 0$

3. *A critical value* c_α *with* $\alpha \in (0,1)$

Conditional on almost all sample paths $Y_{t,t\leq T}$

1. *Under* H_0*:* $\underset{T,B\rightarrow\infty}{p\lim}\,\mathbb{P}^\star(T\mathcal{W}^\star(\theta_p,\Theta_s) \leq c_\alpha | Y_{t,t\leq T}) = \mathbb{P}(T\mathcal{W}(\theta_p,\Theta_s) \leq c_\alpha) = 1-\alpha$

2. *Under* H_1*:* $\underset{T,B\rightarrow\infty}{p\lim}\,\mathbb{P}^\star(T\mathcal{W}^\star(\theta_p,\Theta_s) \leq c_\alpha | Y_{t,t\leq T}) = \mathbb{P}(T\mathcal{W}(\theta_p,\Theta_s) \leq c_\alpha) = 0$

Proof. See Appendix, Properties of Wald Test □

In the Appendix we show that a normal Central Limit Theorem holds, and therefore that the nonparametric bootstrap procedure we describe is valid. We also illustrate using an example that the bootstrap distribution coincides with the asymptotic distribution.

2.7 Estimating the Role of Financial Frictions in Spain

The benchmark economy we use features *some* frictions and since it is standard, we will directly introduce the log-linearized conditions. We consider a small open economy with capital accumulation, along the lines of (Smets and Wouters 2007) and (Gali and Monacelli 2005). There are households, intermediate good firms, final good firms, government expenditure, and the foreign sector, which is composed by infinitesimal symmetric economies.

The type of frictions we allow in the baseline are those we do not have sufficiently informative survey data to implement our methodology. Thus, we keep the parametric Calvo friction in the wage setting by labor unions and in the price setting behavior of firms. However, we remove capital adjustment costs, and therefore Tobin's q becomes constant. This implies that the arbitrage condition between capital and bonds has no dynamics. All other frictions are going to be semi-parametrically characterized. In what follows variables with $*$ denote the "rest of the world", y_t is real output, c_t is consumption, i_t investment, q_t the value of capital, k_t is productive capital, k_t^s capital services, z_t is capital utilization, μ_t^p is the price markup, π_t is domestic inflation, π_{cpi} is CPI inflation, r_t^k is the rental rate of capital, w_t is the real wage and r_t is the interest rate.

$$
\begin{aligned}
y_t &= c_y c_t + i_y i_t + z_y z_t + nx_y s_t + \varepsilon_t^g \\
c_t &= c_1 \mathbb{E}_t c_{t+1} + c_2(l_t - \mathbb{E} l_{t+1}) - c_3(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \\
\mathbb{E}_t r_{t+1}^k &= r_1(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \\
y_t &= \phi_p(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^\alpha) \\
c_t &= c_t^* + \frac{1}{\sigma_a} s_t \\
k_t^s &= k_{t-1} + z_t \\
z_t &= z_1 r_t^k \\
k_t &= k_1 k_{t-1} + (1-k_1) i_t \\
\mu_t^p &= \alpha(k_t^s - l_t) + \varepsilon_t^\alpha - w_t \\
\pi_t &= \pi_1 \pi_{t-1} + \pi_2 \mathbb{E}_t \pi_{t+1} - \varepsilon_t^p \\
\pi_{cpi} &= \pi_t + \nu \Delta s_t \\
r_t^k &= -(k_t - l_t) + w_t \\
\mu_t^w &= w_t - \sigma_l l_t + c_t \\
w_t &= w_1 w_{t-1} + (1-w_1)(\mathbb{E}_t w_{t+1} + \mathbb{E}\pi_{t+1}) - w_2 \pi_t + w_3 \pi_{t-1} - w_4 \mu_t^w + \varepsilon_t^w \\
r_t &= \rho r_{t-1} + (1-\rho)[r_\pi \pi_{cpi,t} + r_y(y_t - y_t^p)] + r_{\Delta Y}[(y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)] + \varepsilon_t^r
\end{aligned}
$$

Let X_1^o denote the vector of variables that enter the moment equalities and X_2^o the vector of variables used in the moment inequalities. Model predictions are denoted with superscript 'm'. Let vector of instruments be Z. The conditions we use are:

$$
\begin{aligned}
\mathbb{E}((X_{1,t}^o - X_{1,t}^m) \otimes Z_t) &= 0 \\
\mathbb{E}((X_{2,t}^o - X_{2,t}^m) \otimes Z_t) &\leq 0
\end{aligned}
$$

We estimate the model using a modification of the MCMC procedure of (Liao and Jiang 2010) with uniform priors. The Appendix discusses sufficient conditions to estimate the identified set using MCMC. [14] The real observables used are in per-capita terms and the

[14]One of the conditions that needs to be satisfied is that the the moment conditions we use, when multiplied by $T^{\frac{1}{2}}$ should satisfy a Central Limit Theorem. While this is straightforward (under assumption **SCI**) for the robust

variables employed in estimation are Non Government Consumption expenditure (C), Hours (H), Inflation (π), Investment (I), Gross Domestic Product (Y), Wages (W), and the EONIA rate (R). As instruments we use realized values of all the variables, lagged one period. We estimate the model using two different subsets of survey data from Spain, collected by the European Commission[15]. Our sample period covers 1999Q_1 to 2013Q_4. Detailed information on this survey data can be found at: `http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm`

In the first set of survey responses, we include responses to quarterly questions 1,2 and 11 from the Consumer Survey, that relate to the financial position of the household. We plot the time series in the Appendix. Credit constraints imply negative distortions to household consumption, and given that hours worked are complementary to consumption, they also imply negative distortions to labor supply and output. We therefore choose $X^o_{1,t} \equiv (W,\pi,R,I)$ and $X^o_{2,t} \equiv (C,H,Y)$. In the second case we use business survey data, in particular questions 8F4 and 8F6, relating to capital adjustment costs and financial constraints to production capacity. For the case of capital adjustment costs, we have restrictions on Y and I similar to those of Example 5. Financial constraints to productive capacity imply similar restrictions and lead to lower aggregate investment and output. We therefore choose $X^o_{1,t} \equiv (C,H,W,\pi,R)$ and $X^o_{2,t} \equiv (Y,I)$.

Some Preliminary Results

For each specification, we obtain estimates of the identified set of wedges, that is, a range for the wedge in each observable that is consistent with survey data and conduct a simple policy exercise. As in statistical decision theory, we specify a decision rule that maps a non singleton set of models (data generating processes) to a single action. Thus, a risk function $\mathcal{R}(Y,d,\mathcal{P}_\theta)$ is chosen and the policy maker chooses the decision $d \in D$ to minimize $\mathcal{R}(Y,d,\mathcal{P}_\theta)$. Because there is multiplicity of $\mathcal{P}_\theta$, even after survey data are used, we need a selection rule. For this application we consider a min-max criteria, $\min_{d\in D}\max_{\theta\in\Theta}\mathcal{R}(Y,d,\mathcal{P}_\theta)$. We choose the combination of $\theta \in \Theta_I$ for which the average (over the observed variables) estimated friction $\lambda(\theta)$ is highest. In Figure 2.3, we plot estimates of consumption wedges consistent with the two different sets of

moment restrictions, we need differentiability in mean for the case of moment restrictions coming from survey data, due to the presence of the non differentiable indicator function.

[15] I thank Fabio Canova for providing the dataset.

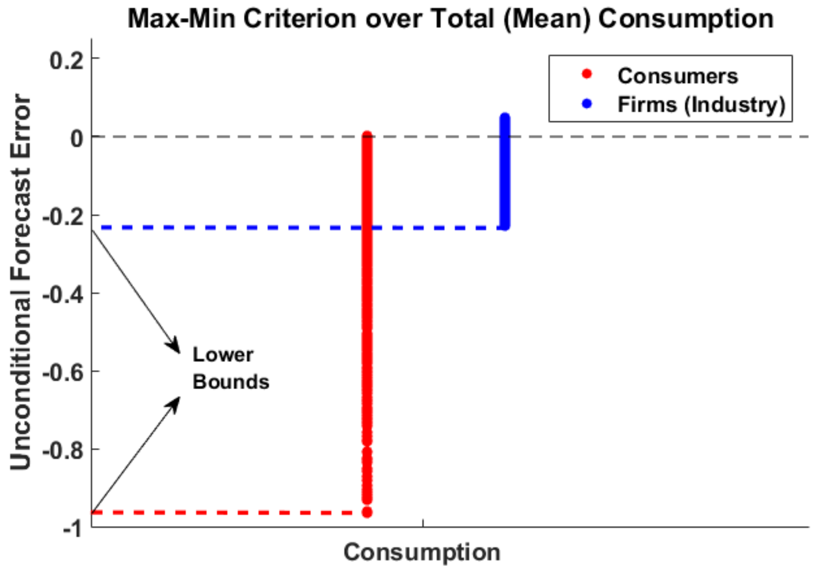


Figure 2.3: Lower bounds to distortions due to different frictions

survey data and the corresponding lower bounds. We plot the estimates of the other variables in the Appendix.

The red (blue) line represents the set of estimates of the distortions to aggregate consumption that are consistent with survey data from consumers (firms). Our pessimistic decision maker chooses a distortion to consumption that is higher in the case of household credit constraints than in the case of financial frictions to firms. While the sign of the red area is determined by the restrictions we impose, the blue set is not since consumption is unrestricted when estimating the model. Nevertheless, because the lower bound to the distortions in output is higher than the lower bound to distortions in investment, and foreign consumption is exogenous, distortions to consumption must also be negative (blue set). In addition, the distortions to output in the worst case scenario are higher if we only take into account financial constraints to firms. This is not surprising, as the share of firms reporting financial frictions in the survey data have been substantially high, especially after 2007. For the other unrestricted variables, we identify negative distortions coming from financial constraints to production in hours worked and negative distortions to investment coming from financial constraints to households.

Given the identified worst case scenario distortions, we can compute a counter-factual conditional path of the optimal interest rate. We restrict attention to the class of Taylor rules of the type: $r_t = \rho r_{t-1} + (1-\rho)[\mu_r + r_\pi \pi_{cpi,t} + r_y(y_t - y_t^p)] + r_{\Delta Y}[(y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)]$ where the objective function is the weighted sum of the variance of the log-deviations from the frictionless steady state of inflation, output and output growth. We plot the interest rate path by plugging in the optimal coefficients and the observed values of the variables. Optimal coefficients are obtained from estimates consistent with consumer survey data or firm survey data. We compare the path we obtain to the path constructed by estimates of the Taylor rule coefficients in (Smets and Wouters 2007) and the realized rate (the marginal lending facility rate). We plot in figures 2.4 and 2.5 the results.

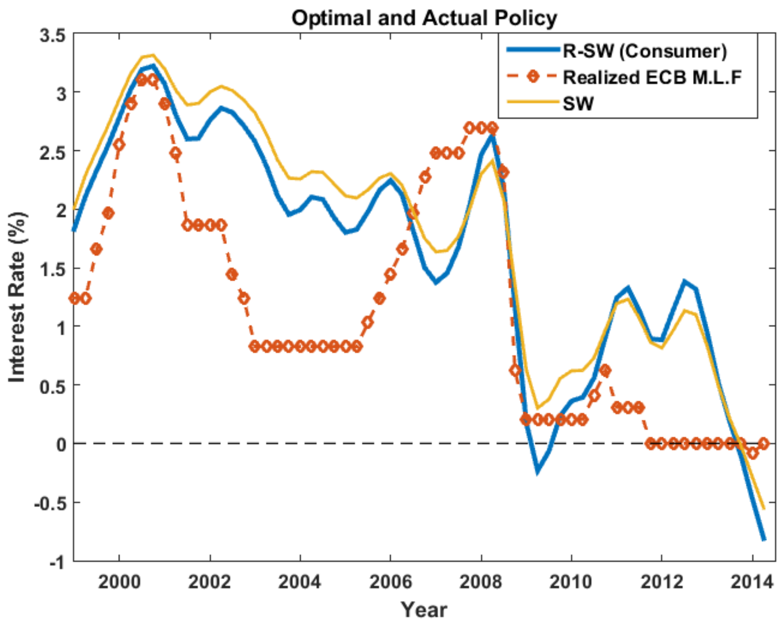


Figure 2.4: Optimal Empirical Taylor Rules - Consumer Data

Figure 2.4, constructed with consumer survey data, indicates that the rate should have been higher during the great moderation period up to the financial crisis. This coincides with the widespread perception that pre-crisis interest rates were too low in some Eurozone countries and this led to the excessive private sector borrowing. Figure 2.5, however, gives a different result: the counterfactual rate should have been lower than the one observed during the same period, because of the relatively high percentage of firms reporting financial constraints even before the crisis. Since this proportion increased dramatically after 2010 our estimate of the optimal interest rate path may be biased downwards. From 2011 and 2013, both Taylor rules do not describe well the path of policy rate mainly due to the fact that inflation recovered from the 2008 - 2009 sharp fall. The persistent fall of inflation after 2013 explains why the counterfactual rate falls into the negative territory.

While we have separated the two sets of survey data for illustration purposes, one could

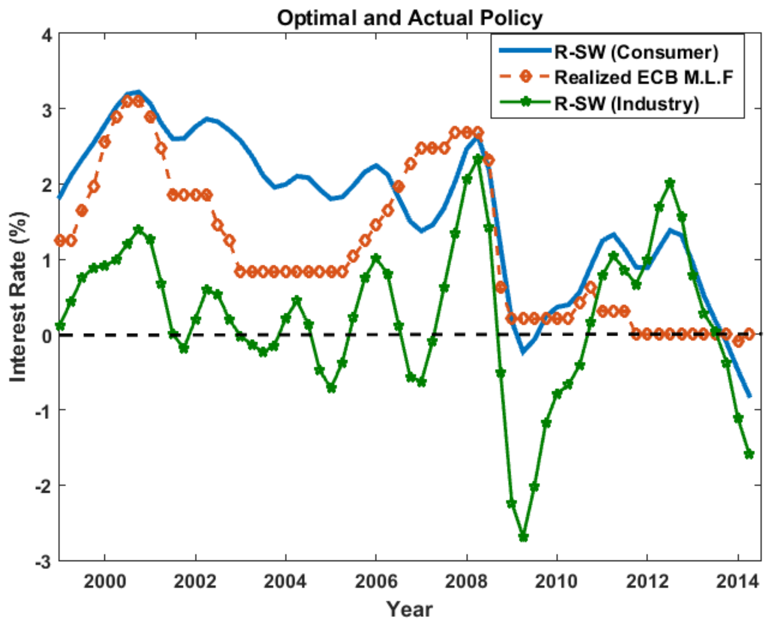


Figure 2.5: Optimal Empirical Taylor Rules - Industrial Data

and should include both of them in the estimation procedure. It is important to stress that our results have nothing to say about regional biases - we need to conduct similar exercises for other countries to reach that conclusion. Nevertheless, the information we have produced is useful since optimal policy in a currency area with heterogeneous frictions implies a monetary policy rule where the weights reflect the degree of frictions present in each country (see for example (Benigno 2004)).

2.8 Conclusion

In this paper we propose a new inferential methodology which is robust to misspecification of the mechanism generating frictions in dynamic stochastic economies. We characterize wedges in equilibrium conditions in a way which is consistent with a variety of generating mechanisms and show how to translate restrictions on the sign of the conditional mean of a wedge in the equilibrium conditions into restrictions on the sign of the conditional mean of the distortion (relative to the frictionless model) in an observable variable. We use the latter restrictions to partially identify the parameters of the model, and to obtain a set of admissible economic relationships.

We also show how qualitative survey data can be linked to the expectations of agents and how this link generates an additional set of identifying restrictions on the probability of observing a distortion in a variable. We state conditions under which the additional restrictions lead to a sharper identified set. We exploit this result to validate parametrically specified models of frictions, propose a weighted Wald statistic, derive its large large sample properties and suggest a bootstrap procedure to compute the critical values.

We apply our methodology to estimate the distortions in the Spanish economy due to financial frictions using an small open economy version of (Smets and Wouters 2007) model and qualitative survey data collected by the European Commission on the financial constraints of the agents. We identify the model generating the maximum distortion to observable variables that is consistent with survey data. We compute optimal policy under the identified worst case model and find that while with survey data from households, the interest rate before the financial crisis was too low relative to the optimal one, the opposite is true when survey data from firms are employed.

In general, our work shows that adopting a robust approach to inference and using the information present in surveys is a fruitful way of dealing with lack of knowledge about the exact mechanisms generating frictions.

One direction where the work can be extended is the following. Throughout the analysis we have focused on the representative agent approximation to the underlying heterogeneous agent economy. We have explicitly considered frictions at the individual level, which when aggregated, produce deviations from the frictionless representative agent world. The effect of heterogeneity is captured by the distribution of the individual state variables. Qualitative Survey data are informative on features of this distribution. Thus, while acknowledging heterogeneity, we use as a benchmark the representative agent model, because it is easy to solve.

(Buera and Moll 2015) have recently shown that the distortions generated in an aggregate equilibrium condition can depend on the type of heterogeneity present in the economy. Our approach is robust to this criticism. The methodology does not rest on *observing* residuals from representative agent frictionless economies - we just impose moment inequality restrictions theoretically motivated by deviations from the frictionless economy. In addition, if heterogeneity has additional implications for the form of these restrictions, they can be taken on board. In addition, we impose weak moment inequalities. This is important when heterogeneous distortions in decision rules cancel out. Finally, we impose restrictions implied by μ_t on all of the variables and thus take general equilibrium effects into account. Nevertheless, future work could focus on investigating the robustness of our methodology in environments where some heterogeneity is ignored when imposing the identifying restrictions.

2.9 Appendix A

Proof. **of Proposition 4**.

Recall the representation for the model with frictions, that is,

$$G(\theta_1,0)X_t = F(\theta_1,0)\mathbb{E}_t(X_{t+1}|X_t) + L(\theta_1,0)Z_t + \tilde{\mu}_t$$

Plugging in the candidate distorted decision rule: $X_t^* = X_t^{f,RE} + \tilde{\lambda}_t$ and using that $\underset{n_x\times n_x}{F(\theta_1,0)}\underset{n_x\times n_x}{P^*(\theta_1,0)} + \underset{n_x\times n_x}{G^*(\theta_1,0)} = 0$ and $(\underset{n_z\times n_z}{R(\theta_1,0)^T} \otimes F(\theta_1,0) + I_z \otimes (F(\theta_1,0)P^*(\theta_1,0) + G^*(\theta_1,0)))vec(Q(\theta_1,0)) = -vec(L(\theta_1,0))$ we have the following condition:

$$\mathbb{E}_t(F(\theta_1,0)\lambda_{t+1} - G(\theta_1,0)\lambda_t + \tilde{\mu}_t) = 0$$

Note that in this proposition we let the econometician's model variables (observables Y_t and unobservables Z_t coincide with a proper subset of X_t and Z_t. That is, setting $\theta_2 = 0$ essentially eliminates some of the elements of (X_t, Z_t). Furthermore, in the proposition we state that $\lambda_t = \lambda_{t-1}\Gamma + \nu_t$ for some real-valued $\Gamma \neq \mathbf{0}$. Substituting for λ_t we get the condition stated in Proposition 3.7, that is:

$$\mathbb{E}_t(F(\theta_1,0)\Gamma - G(\theta_1,0))\lambda_t + \mu_t) = 0 \tag{2.24}$$

To motivate the assumption on the random variable λ_t notice that, the condition above essentially links the conditional mean of λ_t with that of μ_t. Since μ_t is by construction a linear function of X_t, then μ_t is correlated with subsets of the information set of the agent, $\sigma(\mathcal{F}_t)$. Then, there exists a projection operator $\mathbf{P}$ such that for any element $H_t \in \mathcal{F}_t$, $\mu_t = \mathbf{P}H_t + \mathbf{P}_\perp H_t = \mathbf{P}H_t + v_t$. Projecting μ_{t+1} on μ_t:

$$\begin{aligned}\mathbf{P}(\mu_{t+1}|\mu_t) &= \mathbf{P}(\mathbf{P}H_{t+1} + v_{t+1}|\mathbf{P}H_t + v_t)\\ &= \mathbf{P}(\mathbf{P}H_{t+1} + v_{t+1}|\mathbf{P}H_t) + \mathbf{P}(\mathbf{P}H_{t+1} + v_{t+1}|v_t)\\ &= \mathbf{P}(H_{t+1}|H_t)\end{aligned}$$

This implies that, there exists a real valued matrix $\tilde{\Gamma}$ such that $\mu_{t+1} = \tilde{\Gamma}\mu_t + u_{t+1}$. Substituting in the condition 9.1 and collecting all the errors in w_{t+1} (we can do this as they are not uniquely defined) we have that:

$$(F(\theta_1,0)\Gamma - G(\theta_1,0))\lambda_{t+1} \quad = \quad \tilde{\Gamma}(F(\theta_1,0)\Gamma - G(\theta_1,0))\lambda_t + w_{t+1}$$

Denoting by $\tilde{C}$ the generalized inverse of $C := F(\theta_1,0)\Gamma - G(\theta_1,0)$ we have that $\lambda_{t+1} = \tilde{\Gamma}\lambda_t + C\tilde{C}w_{t+1}$. Comparing with the proposed representation for λ_t we have that $\Gamma = \tilde{\Gamma}$ and a non uniquely defined ν_t that nevertheless satisfies $\mathbb{E}(\nu_t) = 0$

Note: We assume that the system can be casted in the expectational form to which we apply the method of undetermined coefficients. We could use more elaborate methods like a canonical or Schur decomposition, so that we explicitly obtain forward and backward solutions. This would complicate the argument without noticeable gains in the intuition why we have an incomplete decision rule. Moreover, the existence of such a rule would require arguments similar to the non zero determinant stability conditions for $\Gamma(\theta)$ and μ_t. □

2.9.1 Characterization of $\mathcal{M}_t$

We restate the optimization problem for a general distance $d(\tilde{\mathcal{M}})$

$$\begin{array}{ll} \max\limits_{\mathcal{M}} & -d(\tilde{\mathcal{M}}) \\ \text{subject to} & \mathbf{1^T}\tilde{\mathcal{M}} = 0 \quad (\lambda_1) \\ & \tilde{\mathcal{M}}^{\mathbf{T}} q_j(Y;\theta) + \mathbf{1^T} q_j(Y;\theta) = 0, \quad j = 1,..p \quad ,(\lambda_{2,j}) \\ & \tilde{\mathcal{M}}^{\mathbf{T}} q_j(Y;\theta) + \left[\mathbf{1^T} q_j(Y;\theta)\right]_+ = 0, \quad j = p+1,..r \\ & \mathcal{M} \geq \mathbf{0} \quad (\lambda_{3,t}, \forall t \in (1..T) \end{array}$$

where $q(Y;\theta)$ is the matrix containing the moment functions and $(\lambda_1,\lambda_2,\lambda_3)$ the corresponding Lagrange multiplier vectors. The first constraint imposes unit expectation while the last constraint imposes non negativity of $\mathcal{M}$. The rest of the constraints impose the moment equalities and inequalities e.g. $[x]_+ \equiv \max(x,0)$

Denote by $\tilde{d}(\mathcal{M}_t)$ as the inverse function of $d(\mathcal{M}_t)$. The Kuhn Tucker first order necessary conditions are the following

$$\begin{aligned}
\mathcal{M}^* &= \mathbf{q}(\mathbf{Y};\theta)\lambda_2 + \lambda_1 + \lambda_3 \\
\tilde{d}(\mathbf{q}(\mathbf{Y};\theta)\lambda_2 + \lambda_1 + \lambda_3)^T \mathbf{q}(\mathbf{Y};\theta) + \mathbf{1}^T \mathbf{q}(\mathbf{Y};\theta) &= 0,\ (\lambda_2) \\
\mathbf{1}^T \tilde{d}(\mathbf{q}(\mathbf{Y};\theta)\lambda_2 + \lambda_1 + \lambda_3) &= 0,\ (\lambda_1) \\
\lambda_{3,t} &\geq 0 \\
\lambda_{3,t}(\tilde{d}(q_t(Y,\theta)\lambda_2 + \lambda_1 + \lambda_{3,t})) &= 0,\ (\lambda_{3,t})
\end{aligned}$$

In the case of chi square distance, that is $d(\tilde{\mathcal{M}}) = \frac{1}{2}\tilde{\mathcal{M}}^{\mathbf{T}}\tilde{\mathcal{M}}$ and ignoring the non - negativity constraint we get an analytical solution. Solving the dual problem and concentrating out the first constraint leads to solutions $(\mathcal{M}^*, \lambda_2^*)$ that satisfy the following system:

$$\begin{bmatrix} I_T & \underset{T\times r}{-(q(Y;\theta) - \bar{q}(Y;\theta))} \\ \underset{r\times T}{q(Y;\theta)^{\mathbf{T}}} & \underset{r\times r}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{T\times 1}{\tilde{\mathcal{M}}} \\ \underset{r\times 1}{\lambda_2} \end{bmatrix} = \begin{bmatrix} \underset{T\times 1}{0} \\ \underset{p\times 1}{-\mathbf{1}^{\mathbf{T}} q_1(Y;\theta)} \\ \underset{(r-p)\times 1}{-\left[\mathbf{1}^{\mathbf{T}} q_2(Y;\theta)\right]_+} \end{bmatrix}$$

We therefore have that for $\tilde{q}(Y;\theta) \equiv q(Y;\theta) - \bar{q}(Y;\theta)$ and $\mathcal{V}(q(Y_t;\theta)) \equiv T^{-1}\tilde{q}(Y;\theta)^{\mathbf{T}} q(Y;\theta)$

$$\begin{bmatrix} \tilde{\mathcal{M}} \\ \underset{r\times 1}{\lambda_2} \end{bmatrix} = \begin{bmatrix} \tilde{q}(Y;\theta)\mathcal{V}(q(Y;\theta))^{-1} \begin{bmatrix} \bar{q}_1(Y;\theta) \\ [\bar{q}_2(Y;\theta)]_+ \end{bmatrix} \\ -\mathcal{V}(q(Y;\theta))^{-1} \begin{bmatrix} \bar{q}_1(Y;\theta) \\ [\bar{q}_2(Y;\theta)]_+ \end{bmatrix} \end{bmatrix}$$

The solution above has been derived ignoring the non negativity constraint. Looking at $\tilde{\mathcal{M}}_t$ the constraint is violated with positive probability, since $\tilde{q}(Y;\theta)$ can take values lower than minus one. Taking into account the non-negativity constraint implies a non analytical solution. There is a variety of algorithms in quadratic optimization to deal with this issue. An alternative way is to use a penalty function that penalizes negative values of $\mathcal{M}_t$. This also typically implies non-closed form solutions. For the adjustment cost example, we re-computed the estimated $\mathcal{M}_t$ and we report it in Appendix B. As is evident, violations of the constraint can be minimal. The existence of a unique function $\tilde{\mathcal{M}} * (\theta)$ implies that the set of models consistent with moment inequalities should have a corresponding one-to-one relation to the identified set, the subset of Θ that satisfies those inequalities: $\theta \in \Theta : \mathbb{E}q_1(Y,\theta) = 0, \mathbb{E}q_2(Y,\theta) \geq 0$. This is made clear in Figure 2.6, which depicts two theory-based moment inequality restrictions on the Euclidean parameter space. The darker area is the identified set, and for the sake of illustration, the point of intersection of the two lines is the combination (θ_1, θ_2) that corresponds to the pseudo-true parameter values of the case of no perturbation $(\mathcal{M}_t = 1)$. The identified set contains the true value, which maps one-to-one to the set of admissible perturbations $\tilde{\mathbb{M}}_2$.

Moreover, choices of objective functional other than $\frac{1}{T}\sum_{t\leq T}\tilde{\mathcal{M}}_t^2$ leads to different sorts of distortions. A general family of distances that can account for non-linearities or non-normalities is the Cressie - Read divergence, of which Chi square is a special case ((Almeida and Garcia 2014, Cressie and Read 1984)). As in the case of non-negative constraints, computing the multipliers might involve numerical optimization. It is also important to stress that for any choice of distance functional, the moment inequality constraints are satisfied. Therefore, the choice of distance functional does not affect the consistency of the parameter estimates.

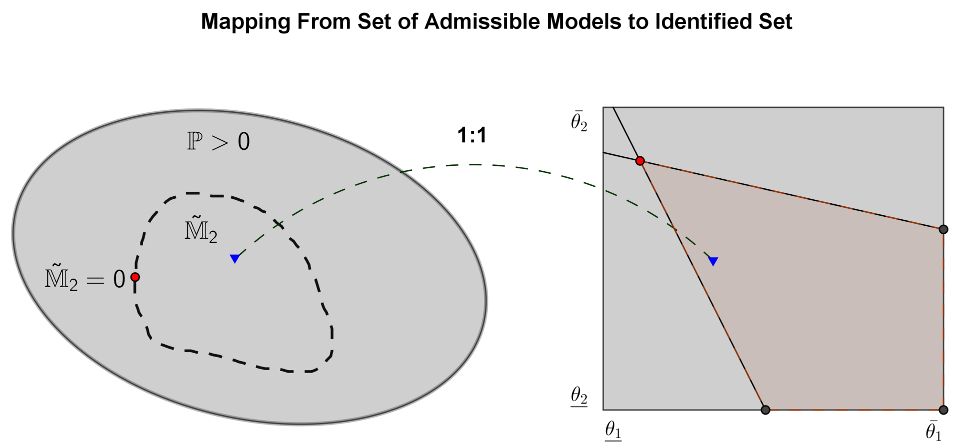


Figure 2.6: Illustration of the mapping from perturbations to the identified set

Proof. **of Proposition 6**: Recall that agent i has the following behavioral equation:

$$G(\theta_1,\theta_2)x_{i,t} = F(\theta_1,\theta_2)\mathcal{E}_{t,i}(x_{i,t+1}|\mathcal{F}_{t-1,i}) + L(\theta_1,\theta_2)z_{i,t} \quad (2.25)$$

$$G(\theta_1,0)x_{i,t} = F(\theta_1,0)\mathbb{E}_t(x_{i,t+1}|\mathcal{F}_{t-1}) + L(\theta_1,0)z_{i,t} + \mu_{i,t} \quad (2.26)$$

Before formally deriving the bounds, we need to establish some facts which will be used in the derivations:

1. **μ_t is a continuous function of the state variables**

   We have already shown that we can move from the model with frictions to the frictionless model (and vice versa) in two ways: Either by setting $\theta_2 = 0$ or by differences in subjective expectations from rational expectations, $\mathcal{E}_{i,t}(X_t|\mathcal{F}_{i,t}) \neq \mathbb{E}(X_t|\mathcal{F}_{t,i}\forall i)$. While the former is quite straightforward, we argue that also the latter can be justified as a change in a component of the model. We have implicitly assumed that the model of the agents, $\mathcal{P}(.|\mathcal{F}_{t,i})$ is absolutely continuous to the Rational Expectations measure $\mathbb{P}(.|\mathcal{F}_{t,i}\forall i)$, and therefore, there exists a Radon Nikodym derivative $\mathcal{M}_{t,i} := \frac{d\mathcal{P}_{i,t}}{d\mathbb{P}_t}$ such that we can for any subset χ of X, $\mathcal{P}_{i,t}(\chi) = \int_{\chi} \mathcal{M}_{t,i}\mathbb{P}_t$. Setting $\mathcal{M}_{t,i} = 1$ for all t, i we get the frictionless model. μ_t varies continuously through θ_2 and $\mathcal{M}_{t,i}$, and is therefore a continuous function of the state variables.

2. **Any probability statement about μ_t translates to a probability statement on λ_t.**

   Since $x^{\star}_{i,t}$ solves the behavioral functional equation of the agent uniquely, there is a map $h : (G,F,L) \to (P,Q)$ which is a continuous bijection, and by the implicit function theorem, any perturbation to the first order conditions (change in (G,F,L)) maps deterministically to perturbations of the solution, (P,Q). Therefore, for every univariate decision variable, $\mathbb{P}(\mu_{i,t} \in [\mathbb{E}_t\underline{\mu}_{i,t}, \mathbb{E}\bar{\mu}_{i,t}]) = \mathbb{P}(h(\lambda_{i,t}) \in [\mathbb{E}_t\underline{\mu}_{i,t}, \mathbb{E}\bar{\mu}_{i,t}]) = \mathbb{P}(\lambda_{i,t} \in [\mathbb{E}_t\underline{\lambda}_{i,t}, \mathbb{E}\bar{\lambda}_{i,t}])$. Same statement holds also for the conditional means of $\lambda_{i,t}$ and $\mu_{i,t}$ given the necessary condition in Proposition 4.

3. **Any probability statement on the subjective conditional expectations translates to a probability statement on $\mu_{i,t}$**

   Recall that we have redefined the variables in the decision problem of the agents such that the enlarged state vector contains also past states, $X_t \equiv (X_t, \tilde{X}_t)$ where $\tilde{X}_t = X_{t-1}$. Observing $\mathcal{E}_{i,t}(x_{i,t+1}) \in B_l$ therefore implies observing either an expectation about the future or the present. Beliefs about the future or the present can therefore be mapped directly to a statement on $\mu_{i,t}$ using the behavioral equation of the agent.

Given the above, we consider the expected value of the statistic $\hat{\mathcal{B}}_{k,t}$. Given the (joint) measure

$\mathbb{P} = \mathbb{P}(t) \times \Lambda(i)$, taking the expectation we have that

$$
\begin{aligned}
\mathbb{E}\hat{\mathcal{B}}_{k,t} &= \sum_{i \le N} w_i \int \int 1(\mathcal{E}_{i,t}(x_{i,t+1}|\mathcal{F}_{t-1,i}) \in B) d(P(\mathcal{F}_t|\mathcal{F}_{t-1}) \times \Lambda(i,t)) \\
&= \sum_{i \le N} w_i \bar{\mathbb{P}}_t(\mathcal{E}_{i,t}(x_{i,t+1}|\mathcal{F}_{t-1,i}) \in B) \\
&= \bar{\mathbb{P}}_t(\mathcal{E}_{i,t}(x_{i,t+1}|\mathcal{F}_{t-1,i}) \in B) \\
&= \bar{\mathbb{P}}_t(\mathbb{E}_t \mu_{i,t} \in [\mathbb{E}_t \underline{\mu}_{i,t}, \mathbb{E}\bar{\mu}_{i,t}]) \\
&= \bar{\mathbb{P}}_t(\mathbb{E}_t \lambda_{i,t} \in [\mathbb{E}_t \underline{\lambda}_{i,t}, \mathbb{E}\bar{\lambda}_{i,t}])
\end{aligned}
$$

In the second to last equality, we use fact **3** and in the last equality we have used fact **2** adapted to conditional means. We now derive the particular result for the positive distortion, $\mathbb{E}_t \mu_{i,t} \ge 0$.

Let $Y_t^o = X_t$ and $Y_t^m = \mathbb{E}X_t$. Using the fact that $(A \Rightarrow B) \Rightarrow \mathbb{P}(A) \le \mathbb{P}(B)$, the independence of the friction and the vector of unexpected shocks, and denoting $\kappa_t \equiv \mathbb{P}_t(c'Z_t \ge 0)$, we have the following bound:

$$
\begin{aligned}
\mathbb{P}_t(\lambda_t + Q(\theta_1,0)Z_t \ge 0) &= \mathbb{P}_t(\int \lambda_i d\Lambda(i,t) + Q(\theta_1,0)Z_t \ge 0) \\
&= \mathbb{P}_t(\int (\mathbb{E}_t\lambda_{i,t} + \lambda_{i,t} - \mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) + Q(\theta_1,0)Z_t \ge 0) \\
&= \mathbb{P}_t(\int (\mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) + \ldots \\
&\quad \ldots + \int (\lambda_{i,t} - \mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) + Q(\theta_1,0)Z_t \ge 0) \\
&= \mathbb{P}_t(\int (\mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) + c'Z_t \ge 0) \\
&\ge \mathbb{P}_t(\int (\mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) \ge 0, c'Z_t \ge 0) \\
&\overset{ind}{=} \mathbb{P}_t(\int (\mathbb{E}_t\lambda_{i,t}) d\Lambda(i,t) \ge 0)\mathbb{P}_t(c'Z_t \ge 0) \\
&\ge \mathbb{P}_t(\cup_i \{\mathbb{E}_t\lambda_i \ge 0\})\kappa_t \\
&\overset{Frechet}{\ge} \max_i \mathbb{P}_t(\mathbb{E}_t\lambda_i \ge 0)\kappa_t \\
&\ge \bar{\mathbb{P}}_t(\mathbb{E}_t\lambda_i \ge 0)\kappa_t \\
&\therefore \\
\mathbb{E}_t(\mathbf{1}(Y_t^o - Y_t^m \ge 0) - \hat{\mathcal{B}}_{k,t}\kappa_t) &\ge 0
\end{aligned}
$$

□

Proof. **of Lemma 8** Given $\mathcal{V}_i(\mathbf{Y}_{t-1})$, the condition in 5.1 can be rewritten as

$$\mathbb{E}(\tilde{Y}_t^o - \underset{n_Y \times n_x}{C(\theta)}\hat{X}_{t|t-1})\phi(\mathbf{Y}_{t-1}) \quad = 0$$

where $\tilde{Y}_t^o \equiv Y_t^o - \mathcal{V}(\mathbf{Y}_{t-1})$. Given **LCI**, the Proposition 2-NS in (Komunjer and Ng 2011) can be applied. Moreover, local identification holds for generic i, and therefore holds for any θ_0 in Θ_I as **LCI** guarantees a unique map from $\mathcal{V}_i$ to θ_0. □

Proof. **of Proposition 9**

Recall that we have the following conditional moment conditions: $q_1(\theta, \mathbf{Y}_{t-1}) \equiv \mathbb{E}(Y_t^o - Y_t^m | \mathbf{Y}_{t-1}) \geq 0$ and $q_2(\theta, \mathbf{Y}_{t-1}) = \mathbb{E}(\hat{B}_t - \mathbf{1}(Y_t^o - Y_t^m > 0) | \mathbf{Y}_{t-1}) \geq 0$.

$$\begin{aligned}
\underset{k \times 1}{q_1(\theta, \mathbf{Y}_{t-1})} &= \mathcal{V}_1(\mathbf{Y}_{t-1}) \in [\underline{\mathcal{V}}(\mathbf{Y}_t)_1, \bar{\mathcal{V}}(\mathbf{Y}_t)_1] \\
\underset{(m-k) \times 1}{q_2(\theta, \mathbf{Y}_{t-1})} &= \mathcal{V}_2(\mathbf{Y}_{t-1}) \in [\underline{\mathcal{V}}(\mathbf{Y}_t)_2, \bar{\mathcal{V}}(\mathbf{Y}_t)_2]
\end{aligned}$$

By choosing suitable instruments Z_t, and any $Z-$measurable function $\phi(.)$, we can construct the following set of unconditional moment conditions,

$$\begin{aligned}
\mathbb{E}(\phi(Z_t)(q_1(\theta, \mathbf{Y}_{t-1}) - \mathcal{V}_1(\mathbf{Y}_{t-1})) &= 0 \\
\mathbb{E}(\phi(Z_t)(q_2(\theta, \mathbf{Y}_{t-1}) - \mathcal{V}_2(\mathbf{Y}_{t-1})) &= 0
\end{aligned}$$

The equations above partially identify the reduced form parameters of the DSGE model. Given a full column rank Jacobian matrix $J(\theta) \equiv \frac{\partial}{\partial \theta}\Lambda(\theta)$, for every observable Y_t we can construct n_θ moment conditions in the first set of moment conditions. Let k be the minimum number of observables in the first set such that $J(\theta)$ is of full column rank. Rewrite the moment conditions such that the first n_θ elements denoted by q, satisfy the rank condition. Given the total number of moment conditions used, m, the rest of the system has $m - n_\theta$ equations. We partition $J(\theta) \equiv (J_r(\theta), J_{n_y - r}(\theta))$ and let $\mathcal{H}$ be an $m \times n_\theta$matrix where $\mathcal{H} \equiv (J_{r \times n_\theta}(\theta)^T, J_{(n_y - r) \times n_\theta}(\theta)^T, \Delta_{\dim q_2 \times n_\theta}(\theta)^T)^T$. Let $m \equiv \mathcal{H} \otimes \phi(Z_t)$ and partition $m \equiv (m_\alpha^T, m_\beta^T)^T$ where m_α contains the first r elements. . Since $m > n_\theta$, and given a general weighting matrix

W, we have the following first order condition:

$$\begin{aligned}
\mathbb{E}(W_{\alpha}^{\frac{1}{2}T} m_{\alpha} + W_{\beta}^{\frac{1}{2}T} m_{\beta} - V(Z_t)\phi(Z_t)) &= 0 \\
\mathbb{E}(W_{\alpha}^{\frac{1}{2}T} m_{\alpha} + W_{\beta}^{\frac{1}{2}T} m_{\beta} - \underset{n_{\theta}\times 1}{U}) &= 0
\end{aligned}$$

This is a projection of m on a lower dimensional subspace. Since W is an arbitrary matrix, and (m_{α}, m_{β}) are possibly correlated, we reproject the sum onto the space spanned by $W_{\alpha}^{\frac{1}{2}T}\mathbf{m}_{\alpha}$. Define $Q_{\alpha} := W_{\alpha}^{\frac{1}{2}T}\mathbf{m}_{\alpha}(\mathbf{m}_{\alpha}^T W_{\alpha}^T \mathbf{m}_{\alpha})^{-1}\mathbf{m}_{\alpha}^T W_{\alpha}^{\frac{1}{2}}$ the projection and $Q_{\alpha}^{\perp}$ the orthogonal projection. Since the original sum satisfies the moment condition, then the two orthogonal complements will also satisfy it: Therefore,

$$\begin{aligned}
Q_{\alpha}\left(W_{\alpha}^{\frac{1}{2}T}\mathbf{m}_{\alpha} + W_{\beta}^{\frac{1}{2}T}\mathbf{m}_{\beta} - U\right) &= W_{\alpha}^{\frac{1}{2}T}\mathbf{m}_{\alpha} + Q_{\alpha}\left(W_{\beta}^{\frac{1}{2}T}\mathbf{m}_{\beta} - U\right) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
Q_{\alpha}^{\perp}\left(W_{\alpha}^{\frac{1}{2}T}\mathbf{m}_{\alpha} + W_{\beta}^{\frac{1}{2}T}\mathbf{m}_{\beta} - U\right) &= Q_{\alpha}^{\perp}\left(W_{\beta}^{\frac{1}{2}T}\mathbf{m}_{\beta} - U\right) \\
&= 0
\end{aligned}$$

where $U \in [W_{\alpha}^{\frac{1}{2}T}\underline{\mathcal{V}}(\mathbf{Y}_t)_{\alpha} + W_{\beta}^{\frac{1}{2}T}\underline{\mathcal{V}}(\mathbf{Y}_t)_{\beta},\quad W_{\alpha}^{\frac{1}{2}T}\bar{\mathcal{V}}(\mathbf{Y}_t)_{\alpha} + W_{\beta}^{\frac{1}{2}T}\bar{\mathcal{V}}(\mathbf{Y}_t)_{\beta}] \otimes \phi(\mathbf{Y}_t)$. Since we are interested in the additional information provided by the second set of restrictions, without loss of generality the first set of restrictions identifies a one to one mapping from U to Θ_I. The second set of restrictions is independent of m_{α} by construction and imposes further restrictions on the domain of variation of U which are implied by the information contained in m_{β} that is orthogonal to m_{α}. By necessity, this implies further restrictions on Θ_I. The same logic has been followed by (Bontemps, Magnac, and Maurin 2012) in the instrumental variable context.

This proof focuses on general moment conditions. To appreciate why the Proposition requires that $\hat{B} \to_p B > 0$ notice that for $B = 0$, $\underline{\mathcal{V}}(\mathbf{Y}_t)_{\beta} = 0$ while the upper bound remains to be $\bar{\mathcal{V}}(\mathbf{Y}_t)_{\beta} = +\infty$. Therefore $U \in [W_{\alpha}^{\frac{1}{2}T}\underline{\mathcal{V}}(\mathbf{Y}_t)_{\alpha}, \infty] \otimes \phi(\mathbf{Y}_t)$. Moreover, the second set of restrictions does not longer provide any information as $Q_{\alpha}^{\perp}U \in [-Q_{\alpha}^{\perp}\varepsilon_t, \infty] \otimes \phi(\mathbf{Y}_t)$ and therefore

$\mathbb{E}Q_{\alpha}^{\perp}U \in [0,\infty]$.

$$\begin{aligned} Q_{\alpha}^{\perp}\left(W_{\beta}^{\frac{1}{2}T} m_{\beta} - U\right) &= Q_{\alpha}^{\perp}\left(W_{\beta}^{\frac{1}{2}T} m_{\beta} + \varepsilon\right) \\ &\Leftrightarrow \\ \mathbb{E}Q_{\alpha}^{\perp}\left(W_{\beta}^{\frac{1}{2}T} m_{\beta}\right) &\in [0,\infty) \end{aligned}$$

is consistent with the unrestricted domain of Θ, that is, any $\theta \in \Theta$ for which $\mathbb{P}(Y_t^o - Y_t^m | \mathbf{Y}_{t-1}) \geq 0)$ is defined is admissible. □

Proof. **of Corollary 11** Suppose that Θ_I is a singleton. For this to be true, it must be that the additional moment restrictions actually hold with equality. Looking at the proof of Proposition 6, this only holds if both $\Lambda(i,t)$ has unit mass on one agent and Z_t has dimension zero. □

2.9.2 Properties of Wald Test and the Block Bootstrap

We first specify the general form of the estimated friction, and we then proceed in analyzing the (first order) large sample behavior of the bootstrap. Recall that the estimates of the frictions satisfy the following program for a general measure of distance, $d(\mathcal{M}_t)$. Denote by $\tilde{d}(\mathcal{M}_t)$ as the inverse function of $d(\mathcal{M}_t)$. The corresponding first order conditions are the following, where bold letters imply vector notation:

$$\begin{aligned} \tilde{d}(\mathbf{q}(\mathbf{Y};\theta)\lambda_2 + \lambda_1 + \lambda_3)^T \mathbf{q}(\mathbf{Y};\theta) + \mathbf{1}^T \mathbf{q}(\mathbf{Y};\theta) &= 0,\ (\lambda_2) \\ \mathbf{1}^T \tilde{d}(\mathbf{q}(\mathbf{Y};\theta)\lambda_2 + \lambda_1 + \lambda_3) &= 0,\ (\lambda_1) \\ \lambda_{3,t}(\tilde{d}(q_t(Y,\theta)\lambda_2 + \lambda_1 + \lambda_{3,t})) &= 0,\ (\lambda_{3,t}) \end{aligned}$$

Although different choices of $d(.)$ lead to different estimates of $\mathcal{M}_t$, the first constraint needs to be satisfied. What this implies is that the inner product of $\mathcal{M}$ and $\mathbf{q}(\mathbf{Y};\theta)$ is equal to $\mathbf{1}^T\mathbf{q}(\mathbf{Y};\theta)$. We therefore resort to analyzing the large sample behavior of plug-in estimates of $T^{-\frac{1}{2}}\sum_t q(Y;\theta)$ and its bootstrap counterpart. In the case of plug in estimates we need to take into account the uncertainty coming from pointwise estimates of $\theta \in \Theta_s$. An alternative way is to construct an $\alpha-$ level confidence interval for Θ_s, $\Theta_{\alpha,s}$. We then use bootstrap to characterize the uncertainty arising from computing the "nuisance parameter", $\gamma \equiv T^{-1}\sum_t q(Y;\theta)$, which is part of the solution to the moment inequality problem, $\inf_{\theta\in\Theta}\inf_{\gamma\in\mathbb{R}^+}(\bar{q}(Y;\theta)-\gamma)'W(\bar{q}(Y;\theta)-\gamma)$.

The Wald statistic we have in mind is the following:

$$TW(\theta_p, \Theta_s) = T(\inf_{\theta \in \Theta_s} ||\mathcal{V}^{-\frac{1}{2}}(f(\theta_p) - f(\theta_s))||)^2$$

where $\mathcal{V}$ is a positive definite diagonal matrix whose elements are the individual variance components of $f(\theta_p) - f(\theta_s)$. Under the Null hypothesis, $H_0 : \theta_p \in \Theta_s$ while under the alternative, $H_1 : \theta_p \notin \Theta_s$. Hereafter we derive the asymptotic distribution of the Wald statistic under the null hypothesis. Let us consider the quantity $T^{-\frac{1}{2}} \inf_{\theta \in \hat{\Theta}_s} ||\mathcal{V}^{-\frac{1}{2}}(\hat{f}(\hat{\theta}_p) - \hat{f}(\hat{\theta}_s))||$.

Given that Θ is a connected set and $\hat{\Theta}_s \in \Theta$, then $\hat{\Theta}_s$ is also connected. For any $\theta, \theta' \in cl(\hat{\Theta}_s)$, $d(\theta, \theta') < \varepsilon$ for arbitrarily small $\varepsilon > 0$. This implies that if $\theta_p \in \Theta_s$ then there exists a $\theta_s \in \hat{\Theta}_s$ such that $||\mathcal{V}^{-\frac{1}{2}}(f(\theta_p) - f(\theta_s))|| < \varepsilon$. Given the moment inequality problem, for every estimating equation, we redefine $\hat{f}(\hat{\theta}_p) = T^{-1}\sum_t q(Y;\theta) - T^{-1}\sum_t \mathbb{E}q(Y;\theta) + \gamma_T = A_T(\theta) + \gamma_T$ where $A_T(\theta) \equiv (A_{T,1}(\theta), A_{T,2}(\theta), ...A_{t,p}(\theta))^T$ and $\gamma_T = (\gamma_{1,T}, \gamma_{2,T}, ...\gamma_{p,T},)^T$. By element-wise mean value expansion around $\theta_l \in \Theta_l, l \in s, p$, $\hat{f}(\hat{\theta}) = \mathbb{E}q(Y_t, \theta_l) + (\hat{f}(\theta_l) - \mathbb{E}q(Y_t, \theta_l)) + D(\tilde{\theta})(\theta - \theta_l))$. Given mild assumptions on the $(2+\delta)$ boundedness of each moment, the second component scaled by $\sqrt{T}$, $\sqrt{T}(\hat{f}(\theta_l) - \mathbb{E}\hat{f}(\theta_l)) \to_d \mathcal{N}(0, \Omega_{1,l})$ while the scaled third component, $D(\tilde{\theta})\sqrt{T}(\theta - \theta_l)) \to_d \mathcal{N}(0, \Omega_{2,l})$. Consequently, for each $j, j \leq p$, $A_{T,j} \to_d \mathcal{N}(0, \Omega^*_j)$. Let $\mathcal{V}(\theta) = AsyVar(T^{\frac{1}{2}} f(\theta_p) - T^{\frac{1}{2}} f(\theta_s))$ and denote by $\mathcal{V}_d(\theta)$ be the matrix containing only the diagonal elements of $\mathcal{V}(\theta)$.

$$T^{\frac{1}{2}}\mathcal{V}_d^{-\frac{1}{2}}(\hat{f}(\hat{\theta}_p) - \hat{f}(\hat{\theta}_s)) = \mathcal{V}_d^{-\frac{1}{2}}T^{\frac{1}{2}}(A_T(\theta_p) - A_T(\theta_s) + \gamma_{T,p} - \gamma_{T,s})$$

Under the Null, $\theta_p \in \Theta_s$, or $f(\theta_p) \in f(\Theta_s)$. Consequently, $\inf_{\theta \in \Theta_s}(\gamma_{T,p} - \gamma_{T,s}) \equiv \inf_{\theta \in \Theta_s}(T^{-1}\sum_t(\mathbb{E}q(Y_t, \theta_p) - \mathbb{E}q(Y_t, \theta_s))) = 0$

$$\begin{aligned} TW(\theta_p, \Theta_s) &= \inf_{f(\theta_s) \in f(\hat{\Theta}_s)} ||\hat{\mathcal{V}_d}^{-\frac{1}{2}}(T^{\frac{1}{2}}A_T(\theta_p) - T^{\frac{1}{2}}A_T(\theta_s) + T^{\frac{1}{2}}(\gamma_{T,p} - \gamma_{T,s}))||^2 \\ &= \inf_{\theta_s \in \hat{\Theta}_s} ||\hat{\mathcal{V}_d}^{-\frac{1}{2}}(T^{\frac{1}{2}}A_T(\theta_p) - T^{\frac{1}{2}}A_T(\theta_s) + T^{\frac{1}{2}}(\gamma_{T,p} - \gamma_{T,s}))||^2 \\ &\xrightarrow{d} ||\sum_{j=1..p} \omega_j \mathcal{N}(0, I_p)||^2 \end{aligned}$$

where $\omega_j = \mathcal{V}_d^{-\frac{1}{2}}\mathcal{V}\mathcal{V}_d^{-\frac{1}{2}}$. Under the alternative, that is when $\theta_p \neq \Theta_s$, and therefore $\inf_{\theta_s \in \Theta_s}(\gamma_{T,p} -$

$\gamma_{T,s}) = O(1)$.

$$\begin{aligned} T\mathcal{W}(\theta_p, \Theta_s) &= \inf_{f(\theta_s)\in f(\hat{\Theta}_s)} ||\hat{\mathcal{V}}^{-\frac{1}{2}} T^{\frac{1}{2}}(A_T(\theta_p) - A_T(\theta_s) + \gamma_{T,p} - \gamma_{T,s})||^2 \\ &= ||\hat{\mathcal{V}}^{-\frac{1}{2}}(T^{\frac{1}{2}} A_T(\theta_p) - T^{\frac{1}{2}} A_T(\theta_s) + \inf_{f(\theta_s)\in f(\hat{\Theta}_s)} T^{\frac{1}{2}}(\gamma_{T,p} - \gamma_{T,s})||^2 \\ &= ||O_p(1) + O_p(T^{\frac{1}{2}})||^2 \\ &= O_p(T) \end{aligned}$$

To get a better approximation to the finite sample distribution of the test statistic, we use a suitable version of bootstrap. Given a bootstrap sample $\{Y^*_{t,l}\}_{t\leq T, l\leq B}$ obtained with a block bootstrap scheme we can compute the wedges to each equation, using the plug-in estimate of θ under the survey-robust case and the full model. We consider the re-centered bootstrapped moments, $(\tilde{f}_1(Y*;\theta), \tilde{f}_2(Y*;\theta)..\tilde{f}_k(Y*;\theta))$ where $\tilde{f}_j(Y*;\theta) \equiv f_j(Y*;\theta) - \bar{f}_j(Y*;\theta)$. We choose to recenter the moments since we deal with an over-identified case, and therefore sample moments, $\bar{q}(Y;\theta)$, are not exactly equal to zero. We obtain critical values by computing the $(1-\alpha)-$quantile of $T\mathcal{W}*(\theta_p, \Theta_s)$. That is, c_α is chosen such that $\mathbb{P}_T(T\mathcal{W}*(\theta_p, \Theta_s) < c_\alpha) = 1-\alpha$. We therefore have that, given the uniform consistency of the bootstrap :

1. Under H_0 : $\underset{T,B\to\infty}{p\lim}\mathbb{P}_T(T\mathcal{W}*(\theta_p, \Theta_s) < c_\alpha | Y_{t,t\leq T}) = \mathbb{P}(T\mathcal{W}(\theta_p, \Theta_s) < c_\alpha) = 1-\alpha$
2. Under H_1 : $\underset{T,B\to\infty}{p\lim}\mathbb{P}_T(T\mathcal{W}*(\theta_p, \Theta_s) < c_\alpha | Y_{t,t\leq T}) = \mathbb{P}(T\mathcal{W}*(\theta_p, \Theta_s) < c_\alpha) = 0$

We illustrate below an example with which we show how the bootstrap behaves in large samples. Small sample distortions is an interesting topic to pursue, but is the subject af another paper. We use a regression based example, which is unrelated to survey data as such, but has the same econometric structure.

Define $C_t \equiv -(D\Omega D')^{-1}D\Omega$ the matrix such that $\hat{\theta}_s - \theta_s = C_t T^{-1}\sum_t q(Y_t, \hat{\theta})$, then we can readily see that partitioning C in the columns that correspond to the initial and supernumerary conditions, $C = (C_1, C_2)$, $T^{\frac{1}{2}}(\hat{\theta}_s - \theta_s) = C_1 T^{-\frac{1}{2}}\sum_t \tilde{q}_1(Y_t, \hat{\theta}) + C_2 T^{-\frac{1}{2}}\sum_t \tilde{q}_2(Y_t, \Xi_t, \hat{\theta})$. The expression for $T^{\frac{1}{2}}(\hat{\theta}_p - \theta_p)$ is $C_1 T^{-\frac{1}{2}}\sum_t \tilde{q}_1(Y_t, \hat{\theta})$. Then, under the Null, $T^{\frac{1}{2}}(\hat{\theta}_s - \theta_s - (\hat{\theta}_p - \theta_p)) = C_2 T^{-\frac{1}{2}}\sum_t \tilde{q}_2(Y_t, \Xi_t, \hat{\theta})$, which satisfies a standard Central Limit Theorem.

Note: We maintain regular inference by dealing with the multiple inequality issue (affecting $\hat{\theta}_s$) using the method proposed by (Chernozhukov, Kocatulum, and Menzel 2012), that is using a single smooth inequality, that approximates the intersection of the multiple inequalities.

More particularly, we use $\mathbb{E}(q(Y_t,\hat{\theta}_s)) \equiv \sum_j \frac{\exp(\iota\mathbb{E}(q_j(Y_t,\hat{\theta}_s)))}{\sum_j \exp(\iota\mathbb{E}(q_j(Y_t,\hat{\theta}_s)))}\mathbb{E}(q_j(Y_t,\hat{\theta}_s)))$ for some constant $\iota > 0$. This is an approximation to $\max_j \mathbb{E}(q_j(Y_t,\hat{\theta}_s))$. The identified set $\Theta_{s,\iota}$ therefore depends on ι. The more well separated $\max_j \mathbb{E}(q_j(Y_t,\hat{\theta}_s))$ is from the rest inequality generating functions the better the approximation. As shown in (Chernozhukov, Kocatulum, and Menzel 2012), in general $\Theta_s \subset \Theta_{s,\iota}$, that is the approximation is conservative. All of our results therefore are directly on $\hat{\Theta}_{s,\iota}$.

Example 7. ***Measurement error in Regressors.*** *Suppose there are two independent measurements of a regressor and the model for the measurement error is $X_{1,t} = X_t^* + \nu_{1,t}$ and $X_{2,t} = X_t^* + \nu_{2,t}$ respectively. Furthermore assume that $\nu_{1,t} \sim \mathcal{N}(0, 0.2^2)$ and $\nu_{2,t} \sim \mathcal{N}(0, \kappa\sigma_{\nu_{1,t}})$ for $\kappa > 1$, $X_t^* \sim \mathcal{N}(0, 0.1)$ and $\varepsilon_t \sim \mathcal{N}(0, 0.1)$. For simplicity (and this is inconsequential for the test), we assume that she uses only one of the measurements. The important assumption in this case is that she "knows" all parameters apart from β and that she **mistakenly assumes** that $\sigma_{\nu_{1,t}} = 0.5$. She uses Simulated Maximum likelihood to estimate β_m where $Y_t = 0.2 + \beta X_t + \varepsilon_t$. A robust approach would be to be agnostic about the distribution of the errors and use the well known fact that $B_0 = \{\beta \in B : \beta_{ols} \leq \beta\}$. Using both measurements, this would mean that $B_0 = \{\beta \in B : \beta_{ols,1} \leq \beta \cap \beta_{ols,2} \leq \beta\}$. Since $\beta_{ols,1} > \beta_{ols,2}$ then the identified set is determined by the first measurement which is also used by the econometrician. We therefore test $H_0 : \beta_m \in B_0$.*

To see the equivalence of this test to the test we propose, notice that under the Null $T^{-\frac{1}{2}} \inf_{\beta_0 \in B_0}(X^T(Y - \hat{\beta}_0 X) - X^T(Y - \hat{\beta}_m X)$ is equal to $T^{-\frac{1}{2}} X^T X(\hat{\beta}_m - \beta_m - (\hat{\beta}_0^* - \beta_0^*))$. The residual in this case, $\sum_t \tilde{q}_2(.)$ is equal to $(\frac{X^T X}{T})^{-1} X^T \varepsilon + \beta_0 \frac{X^T X^*}{X^T X} + \frac{\tilde{X}^T \tilde{X}}{T})^{-1} \tilde{X}^T \varepsilon + \beta_0 \frac{\tilde{X}^T X^*}{\tilde{X}^T \tilde{X}}$. We plot below the bootstrap distribution versus the simulated test statistic, which in this case according to our theoretical result is, for $\lambda_{x_1,x_2} \equiv \frac{\sigma_{x_1}^2}{\sigma_{x_2}^2}$, $TW \sim (\beta^2((1-\lambda_{x,x^*})(\lambda_{x,x^*} - 2\lambda_{x,\tilde{x}}) + \lambda_{x,\tilde{x}}(1 - \lambda_{x,\tilde{x}})) + \lambda_{\varepsilon,x}(\lambda_{x,x^*} - \lambda_{x,\tilde{x}})\chi^2(1)$

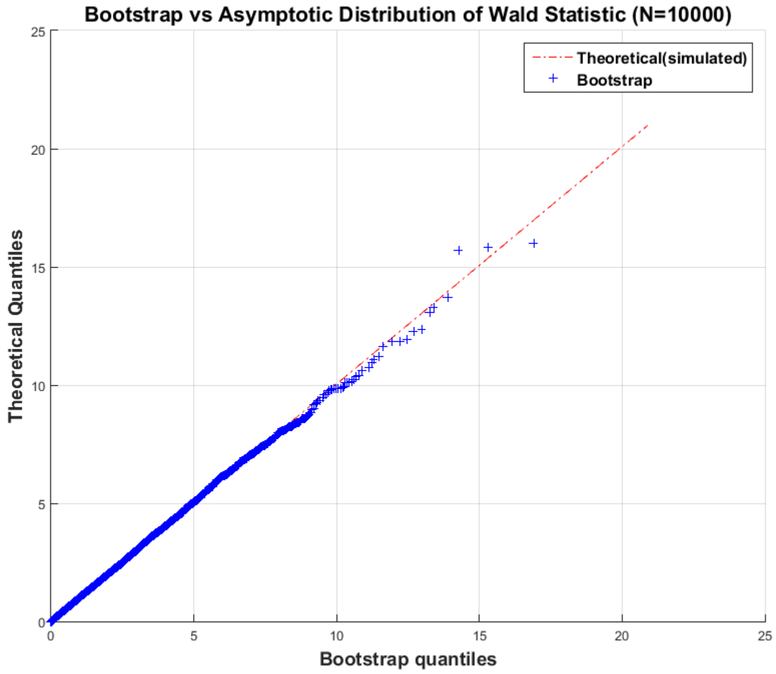


Figure 2.7: Q-Q plot of Bootstrap versus Asymptotic distribution of W^*

2.9.3 Using MCMC to explore the GMM pseudo-likelihood

It has been shown in (Chernozhukov and Hong 2003) that the Generalized Method of Moment (GMM) class of estimators can be easily embedded in a Laplace type of estimation. Although the authors refer to models that are point identified, this limited information approach has been considered also by (Liao and Jiang 2010) in a partially identified case. One of the main assumptions which justify the use of a quasi likelihood approach to GMM is the fact that a scaled (by $n^{\frac{1}{2}}$) moment condition is asymptotically Normal. The resulting pseudo-likelihood cannot be considered as fully characterizing the probability distribution of the data as GMM involves information loss. Nevertheless, the pseudo likelihood obtained can be used to characterize the large sample frequentist properties of θ.

In our case, the moment functions generated by the DSGE model are assumed to be well behaved. Nevertheless, since we also use quantile functions to characterize the restrictions implied by the survey data, we can no longer invoke smoothness assumptions. This does not pose significant difficulty though as we can resort to sufficient stochastic equicontinuity conditions.

These conditions are of the following type:

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{|\theta-\theta'|\le\delta} \frac{n^{\frac{1}{2}}|Q_n(\theta)-Q_n(\theta')-(\mathbb{E}Q_n(\theta)-\mathbb{E}Q_n(\theta')|}{1+n^{\frac{1}{2}}|\theta-\theta'|} > \varepsilon \right\} \quad < \quad \varepsilon$$

In our case it is sufficient to look at $Q_n \equiv \frac{1}{T}\sum_t((\mathbf{1}(Y_t^o - Y_t^m(\theta) \ge 0)\phi(\mathbf{Y}_{t-1}))$ as survey data do not depend on θ. Since we are in the class of linear models, it is the case that $Y_t^m(\theta) = C(\theta)X_{t,t}(\theta)$ where $X_{t,t}$ is $\mathbf{Y}_{t-1}$- measurable. The output of the Kalman filter $(X_{t,t})$ is a linear function of the observable $Y_{t-1}$and we can rewrite $Y_t^m(\theta) = C(\theta)X_{t,t}(\theta) = D(\theta)Y_{t-1}$ and $D(\theta)$ continuously differentiable with respect to θ (given stability conditions, a well behaved $\Sigma_{t,t}$exists). Stochastic equicontinuity of $Q(D)$ implies stochastic equicontinuity w.r.t θ. In order to verify stochastic equicontinuity of Q in the case of dependent data, certain assumptions have to be made about the degree of dependence allowed and the complexity of the function class $\mathscr{M}$ considered. Sufficient conditions are the following ((Andrews 1993, Doukhan, Massart, and Rio 1995)):

1. $\{Y_t : t \ge 1\}$ is a stationary absolutely regular sequence with $\beta(s) \le C\tau^s$ for some $\tau \in (0,1)$.
2. $\mathbb{E}\bar{M}^2(Y_t)\log\bar{M}(Y_t) < \infty$
3. $\log \mathscr{N}_p^B(\varepsilon, \mathscr{M}) \le C(\frac{1}{\varepsilon})^B$ for some $B < \frac{1}{2}$ and $p > 2$.

where $\bar{M}$ is the envelope function of $\mathscr{M}$, that is $\max_i \sup_\theta q_i(Y_t;\theta) \le \bar{M}(Y_t)$ and $\mathscr{N}_p^B(\varepsilon, \mathscr{M})$ is the L^pbracketing cover number.

Assumption 1 is satisfied for $VAR(1)$ processes as long as the innovations have a bounded density with respect to Lebesgue measure and finite $2+\delta$ moments, $(\delta > 0)$. Nevertheless, DSGE's most likely have $VARMA(p,q)$ representations. Moreover, if we let $q_i(Y_t;\theta) = (\mathbf{1}(Y_{i,t}^o - Y_{i,t}^m(\theta) \ge 0)\phi_i(\mathbf{Y}_{t-1}) \in \mathscr{M} \equiv \mathscr{M}_1\mathscr{M}_2$ where subscripts differentiate between the indicator function class and the instrument function class. Then assuming $\mathbb{E}\max_i \sup_\theta(\log\phi_i(\mathbf{Y}_{t-1}))^{2+\delta} < \infty$ for any measurable function of Y_t is sufficient for Assumption 2. Finally, regarding Assumption 3, we can combine the cover numbers of $\mathbf{1}(Y_{i,t}^o - Y_{i,t}^m(\theta) \ge 0)$ and $\phi(\mathbf{Y}_{t-1})$ as in (Andrews 1993). It is well known that the indicator function has a bracketing number that grows at a polynomial rate. Moreover, $\phi(\mathbf{Y}_{t-1})$ is typically not a function of θ, although if we use (Kalman) filtered state estimates, it will be. In the latter case, assuming Lipschitz continuity, compact Θ and sufficiently smooth instrument functions with finite $2+\delta$ is sufficient. More particularly, $\log\mathscr{N}_p^B(\varepsilon,\mathscr{M}) = \log\mathscr{N}_p^B(D\varepsilon,\mathscr{M}_1) + \log\mathscr{N}_p^B(D\varepsilon,\mathscr{M}_2)$ where $\log\mathscr{N}_p^B(D\varepsilon,\mathscr{M}_1) \sim \frac{D}{\varepsilon}$,

$\log \mathcal{N}_p^B(D\varepsilon, \mathcal{M}_2) \sim (\frac{1}{D\varepsilon})^{1/(2+\delta)}$ and $D \sim \frac{1}{\mathbb{E}\max_i \sup_\theta (\log \phi_i(\mathbf{Y}_{t-1}))^{2+\delta}}$ The second bracketing number dominates, and it grows at a rate slower than $(\frac{1}{\varepsilon})^{1/2}$, satisfying therefore $\mathbb{P}-$Donskerness.

Given these conditions are satisfied by the quantile-type of functions we use, then the results of (Chernozhukov and Hong 2003) and (Liao and Jiang 2010) carry through. More particularly, let $S_n(\theta) = n^{\frac{1}{2}} q_n(\theta)'_+ W_n n^{\frac{1}{2}} q_n(\theta)_+$ be the criterion function to be minimized, where $q_n(\theta)$ are the moment functions to be used and $q_n(\theta)_+ \equiv \max(q_n(\theta), 0) = \min(-q_n(\theta), 0)$. Let correspondingly $p_n(\theta)$ be proportional to the pseudo density induced by GMM, $p_n(\theta) \propto \exp(-S_n(\theta)$. Then there exists a $\Delta_n(\theta_0)$ and $J_n(\theta_0)$ such that $S_n(\theta)$ admits a quadratic expansion, that is

$$S_n(\theta) \quad = S_n(\theta_0) + \quad (\theta-\theta_0)'\Delta_n(\theta_0) - \frac{1}{2}(\theta-\theta_0)' n J_n(\theta_0)(\theta-\theta_0) + R_n(\theta)$$

It is important to note that in this context the existence of the terms $\Delta_n(\theta_0)$ and $J_n(\theta_0)$ does not rest on differentiability assumptions for $q_n(\theta)$ but rather on differentiability in mean. This allows us to assume the existence of a CLT on $\Delta_n(\theta_0)$ without smoothness assumptions on $q_n(\theta)$. Furthermore, let me redefine $Q_n(\theta) \equiv q_n(\theta) - b$ where b is t he bias term. That is instead of looking at $\inf_\theta S_n(\theta)$ we can equivalently look at $\inf_{(\theta, b: b \in [0,\infty))} \tilde{S}_n(\theta, b)$ where $\tilde{S}_n(\theta, b) = n^{\frac{1}{2}} Q_n(\theta)' W_n n^{\frac{1}{2}} Q_n(\theta)$. we can therefore assume that the following holds:

$$V(\theta_0, b_0)^{-\frac{1}{2}} \tilde{\Delta}_n(\theta_0, b_0) \quad \xrightarrow{d} \quad N(0, I)$$

Given this result, in (Liao and Jiang 2010) it is shown that the posterior density (in our case, simply the quasi likelihood) drops to zero exponentially on any subset $\delta > 0$ away (in terms of Euclidean distance) from Θ_I^c. Given this exponential rate, we can also estimate the identified region.[16] See (Liao and Jiang 2010) for further details.

We should bare in mind that it can be the case that the credible set estimated might be inside the identified region ((Moon and Schorfheide 2012)). We do not perform inference based

[16]Define $(\Theta_I^c)^{-\delta} = \theta \in \Theta : d(\theta, \Theta_I) \geq \delta$ where $d(x,y)$ is the Euclidean distance. What (Liao and Jiang 2010) show amongst other results is that:

1. $\forall \delta > 0$ and for some $\alpha > 0$,

$$\mathbb{P}(\theta \in (\Theta_I^c)^{-\delta} | \mathbf{Y}) \quad = \quad o_p(\exp(-\alpha n))$$

2. $\forall$ nonempty open sets $\Theta_s \subset \Theta_I$,

$$\liminf_{n\to\infty} \mathbb{P}(\theta \in \Theta_s | \mathbf{Y}) \quad > \quad 0$$

on Markov Chain Monte Carlo, we use it only to get a consistent estimate ofthe identified set. We resort to the Bootstrap to do inference of the type $H_0 : \theta = \theta_0$.

2.10 Appendix B

2.10.1 Computational results for the case of Capital Adjustment Costs

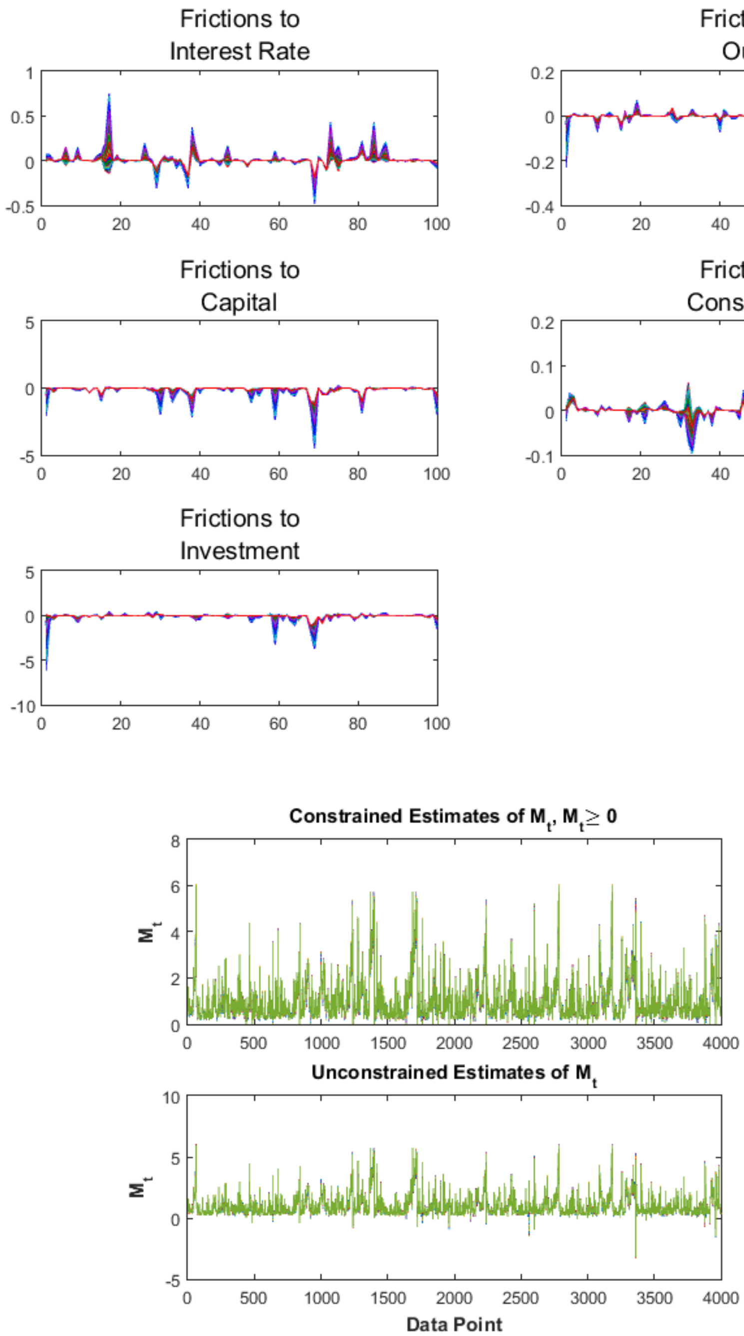


Figure 2.8: (Top) Estimates of $(1-\mathcal{M}_t)\lambda_t$ and (Bottom)Violations of $\mathcal{M}_t \geq 0$ over the whole artificial sample (100 periods)

2.10.2 Graphical Examples of Spanish Survey Data

The data used come from surveys conducted by the local statistical authority under the guidelines of the European Commission. What we present in the graphical evidence comes from the sub-components of the Consumer Survey Index (CSI) and the Industrial Sector Index (ISI). The indices constructed reflect the balances of the designated answers. In particular, the answers are: " better (PP), little better (P) , same (E), little worse (M), lot worse (MM), N don't know. Then, balances are calculated as B = (PP+0.5P)-(0.5M+MM). We plot below aggregated answers to typical questions relating to savings in the household sector and production constraints due to equipment and financial constraints in the industrial sector.

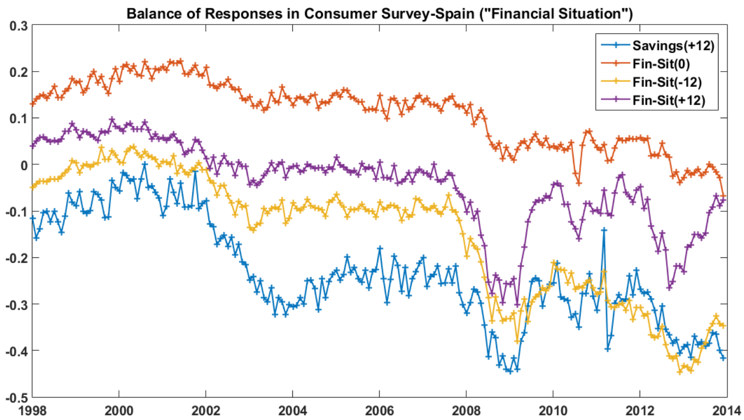


Figure 2.9: Spain

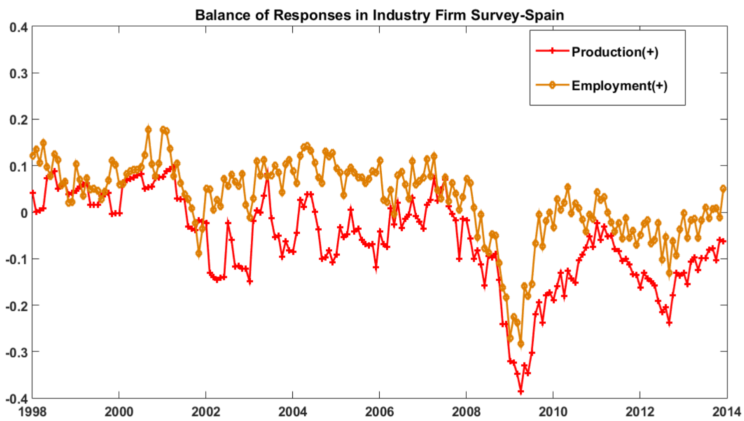


Figure 2.10: Spain

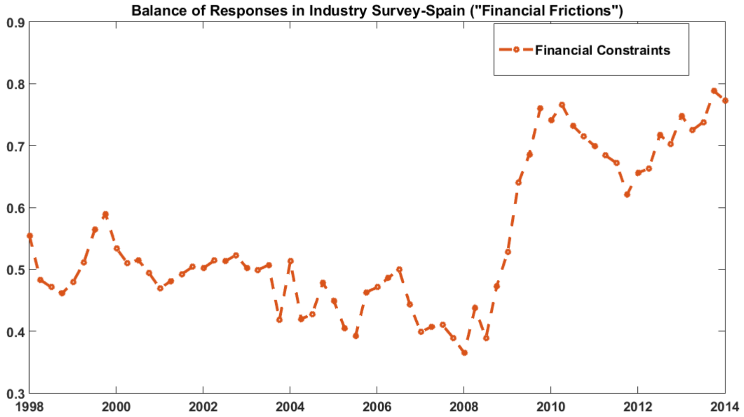


Figure 2.11: Spain

2.10.3 Data Transformation and Filtering

2.10.3.1 Macro Aggregates

We extract the business cycles from Spanish macro aggregates by applying the Christiano - Fitzgerald optimal approximation to the Band-Pass filter for length ranging from 4 to 32 quarters. Figure 1.12 depicts the extracted cycles for each series.

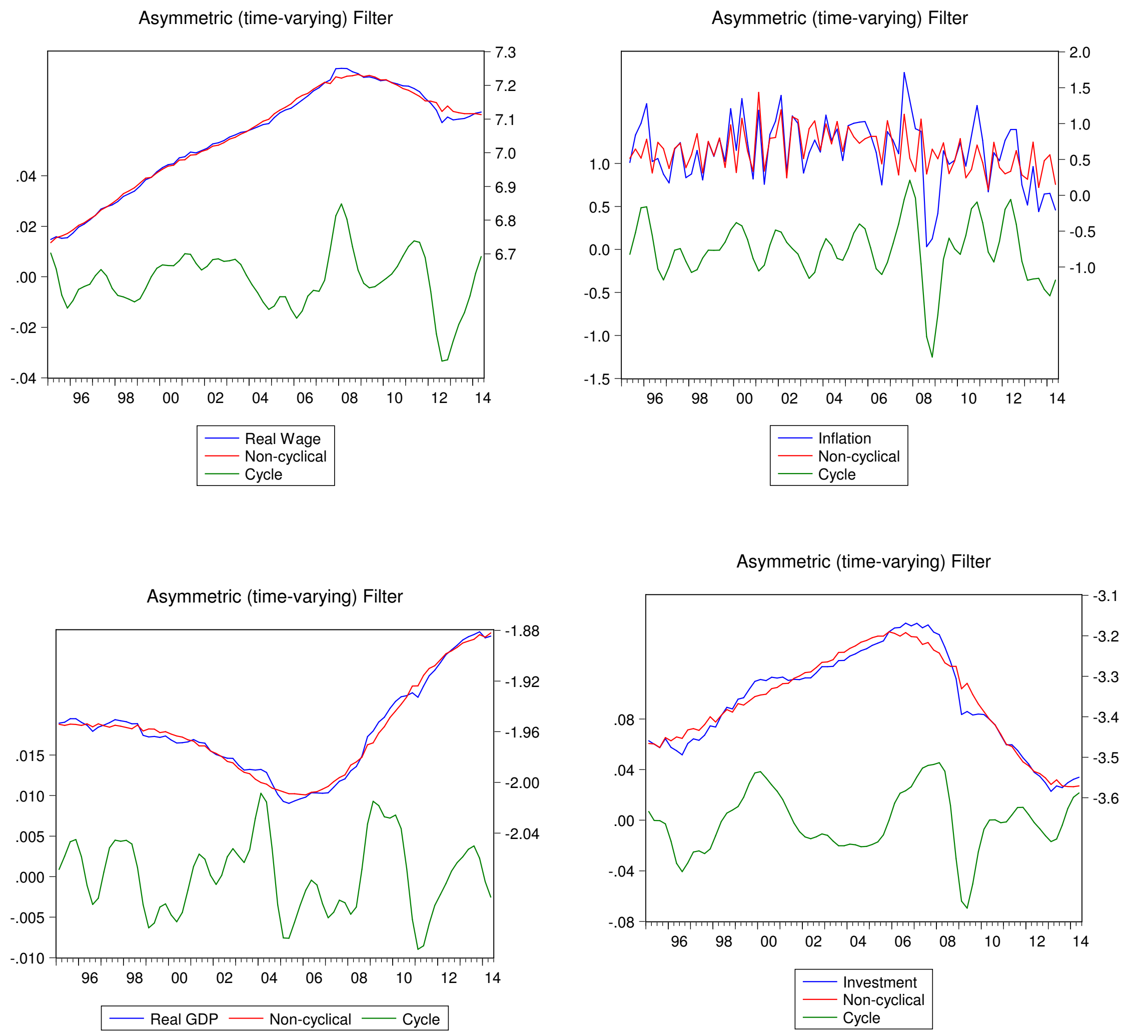


Figure 2.12: Spain

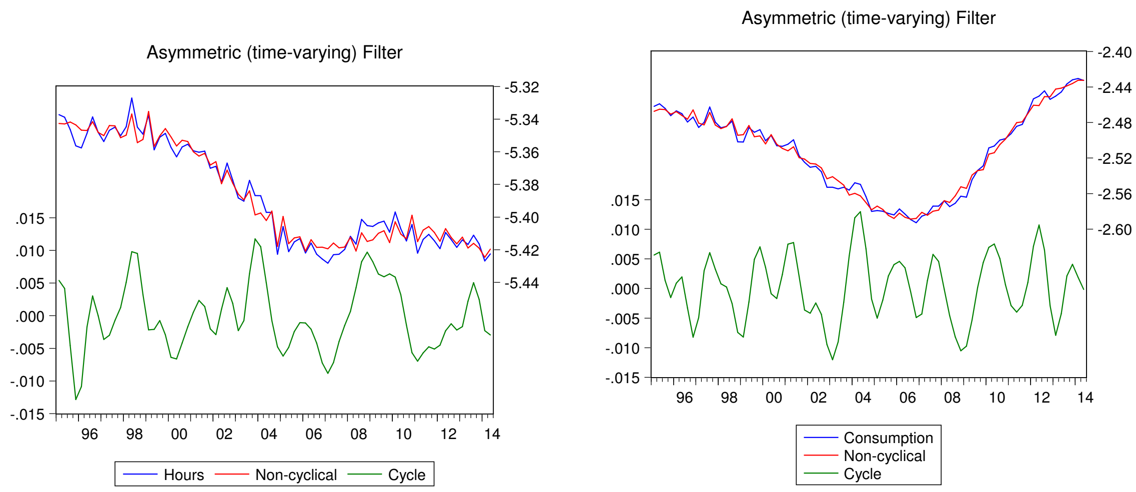


Figure 2.13: Spain

2.10.3.2 **Balance Statistics**

Given that in our theoretical analysis we assumed positive balances that directly relate to a probability statement, we need to make a simple transformation of the balance statistics. We have already defined the probability of an event as a certain partition of the relevant random variable of interest. If this random variable is x, then we partition x in i.e. 5 intervals, namely x_1, x_2, x_3, x_4, x_5. The probability mass p_i in a particular partition will therefore give the probability of this random variable lying in this partition of the support. The balance statistic $B = p_5 + 0.5p_4 - p_2 + 0.5p_1$ is equivalent to judging whether there is more mass above or below the median x, $med(x)$ and truncating the distribution on both ends by the same proportion to avoid extremes. Taking the truncated distribution as the true distribution, then the balance statistic is $B \equiv \mathbb{P}(x > med(x)) - \mathbb{P}(x < med(x))$ which implies that $\mathbb{P}(x > med(x)) = \frac{1+B}{2}$.

2.10.3.3 **Interpretation of $\mathbb{P}(x > med(x))$**

In the easiest case, that one in which agents report constraints, $\mathbb{P}(x > med(x))$ is the probability of having constraints. In more subtle cases, when agents report on the situation being the same, better or worse, $\mathbb{P}(x > med(x))$ is the conditional probability of a transition to a better or worse state. Since decision rules from DSGE models depict decisions conditional on a state variable, then $\mathbb{P}(x > med(x))$ is the conditional probability on the evolution of an endogenous or trivially an exogenous variable, or a combination of both. Such a subtle case arises in our application, in which we have to link consumer qualitative survey data to aggregate implications for borrowing constraints (incomplete markets).

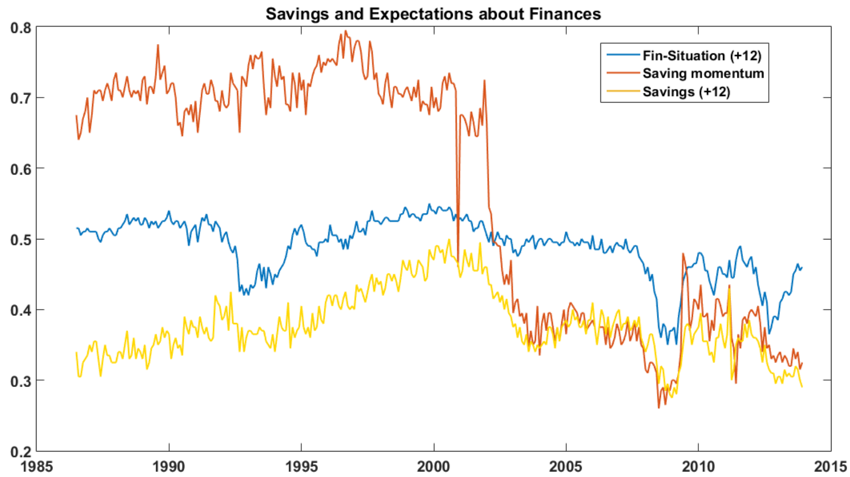


Figure 2.14: Spain

In the figure above we plot the proportion of people who claim that "it is a better moment to save than not" (orange) versus proportion of people who believe that "their savings will improve rather than not" (yellow). What is striking is that the two proportions coincide in the following sense: Before the mid 1990's global recession, roughly the same proportion of people believed that it was a good moment to save and that their savings would deteriorate in the future, which is an indication of behavior consistent with saving to smooth fluctuations in income. Agents expect to run down savings in the future due to an expectation of worsening financial situation. From 1997, more agents expected to save in the future and at the same time believed that it was a good moment to save. The latter reversed during the boom period as housing markets were booming. From 2008, with the outburst of the crisis, the proportion of people believing that it is rather not a good time to save and that of expecting of saving to fall further in the future, co-move (the complements of orange and yellow). This is also consistent with consumption - saving behavior in incomplete markets, as agents use their existing savings to smooth their consumption. We relate the blue series (expected financial situation in 2 months) to "cash on hand", that is assets and income, and the yellow series to expectations about income (labor and capital), since savings crucially depend on the latter in any model of consumption - savings decision. The difference between the blue and the yellow (and orange after 2003) implies that not all agents who believe that their financial situation will improve will necessarily save more in the future. We conclude that the complement of the probability implied by the yellow series is a rough approximation to the probability of having i.e. a negative distortion to the level of consumption relative to the frictionless case as it's behavior over time is consistent with incomplete markets. We nevertheless include the other series i.e. the complement of the blue series, which is a lower bound to the probability level we want to characterize. It is co-moving with the yellow series so it might lower variance at the expense of bias.

We should note that we have not controlled for "breaks" in the survey data series for a few reasons. First, we have a small sample and the cost of dropping observations is quite high. Second, from the data series we use, only the proportion of firms having financial constraints seems to have a "structural break", see Figure 2.11. This might affect to some extent the level of the correlation of this series with the instruments we use, which is what matters for estimation.

2.10.4 Distortions due to different frictions

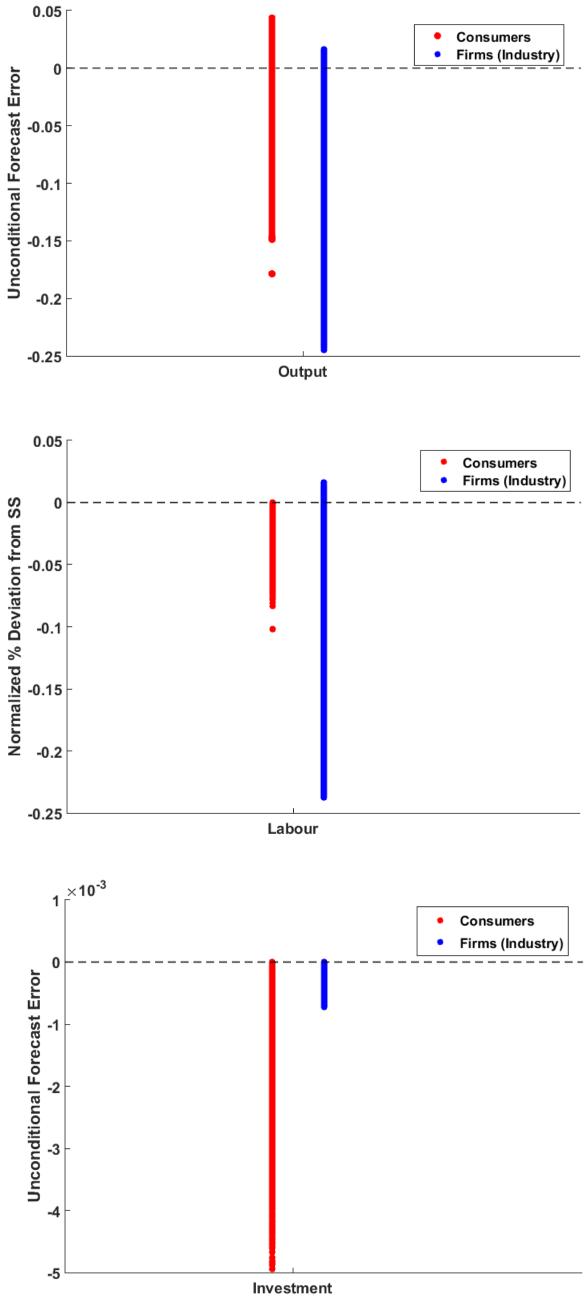


Figure 2.15: Spain

Chapter 3

Estimation and Inference for Incomplete Structural Models using Auxiliary Density Information

3.1 Introduction

The use of estimated Dynamic Stochastic General Equilibrium (DSGE) models has become pervasive in both economic policy and academic 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 macroeconomic theory provides a set of equilibrium conditions, it rarely provides the complete probability distribution of observables, which is necessary to perform full information analysis. 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, approximation and model solution can introduce further uncertainties like loss of identification power (Canova and Sala 2009).

With regard to the form of the solution and types of equilibria considered, although some solutions or 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). On the other hand, having a complete probability distribution is very useful. It enables practitioners to do counterfactual experiments, predictions and therefore policy recommendations.

This paper considers an alternative method for estimating the parameters of a DSGE 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.

This approach can allude to the Bayesian paradigm in the sense that the approximate model serves as a prior, which can nevertheless be data revisable. We provide a decision theoretic framework that rationalizes the estimation method, and develop the corresponding frequentist inference. We limit most of our analysis to the case of finite dimensional φ, although 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, and this is reflected in the first order estimating equations. 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 semi-parametric lower bound for the parameter estimates is attained (see (Chamberlain 1987)). Another contribution of the paper is to show that under misspecification, the estimator can be rewritten in the form of a regularized GEL estimator in which shrinkage is towards the otherwise misspecified conditional density. Misspecification of the density in the form of improper finite dimensional restrictions leads to efficiency gains and therefore an asymptotic bias - variance trade-off.

Moreover, we provide simulation comparisons of the Mean Squared Error (MSE) of the estimator for the case of fixed density misspecification. Well specifying the conditional mean is important to get good estimates in the MSE sense. We also apply the method to the prototypical

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

stochastic growth model.

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), and Generalized Empirical Likelihood (GEL) criteria i.e. (Newey and Smith 2004) in a conditional moment restrictions framework. We depart from this literature by considering a generalized version of exponential tilting, where the form of $f(X|Z,\varphi)$ is parametrically specified.

The paper is organized as follows. In Section 2, we introduce the problem of likelihood recovery and provide a decision theoretic interpretation to the method. Moreover, an economic example for density projections is provided. In Section 3 we provide 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 application, and discusses the computational aspect of the method and the case of non differentiable models. Appendix B contains the proofs.

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 belong to a probability space $(\Omega, \mathcal{F}, P)$. In the paper three different probability measures are used, the true measure P, the base measure F_φ which is indexed by parameters φ and the $H_{(\varphi,\vartheta)}$ measure which is obtained after the information projection. Moreover, these measures are considered absolutely continuous with respect to a dominating measure v on the space where X is defined e.g. if $X \in \mathbb{R}^{n_x}$ then v is a 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, $m_i(X,Z,\vartheta)$ is a general moment $X \otimes Z$ measurable function and $m(X,Z,\vartheta)$ is an $M \times 1$ 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]. $\mathbb{E}_P(.|.)$ is the conditional mathematical expectations operator with respect to measure P. Finally, $\mathbb{V}_P(x)$

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

signifies the variance of variable x under the $P-$ measure while $V_{\tilde{P},s}(x)$ is the second moment of a particular function $\tilde{s}(.)$.

3.2 Incomplete Models, Likelihood and (Non) Revisable Information

In this section we provide a decision theoretic motivation for the methodology by casting the issue of likelihood recovery as a problem that involves the introduction of non data-revisable information. Auxiliary modeling choices like model approximations and choice of particular equilibria entail a loss to the econometrician or policy maker that is non revisable once the selected model is taken to the data.

We can model such a loss by considering the decision problem of a modeler who makes predictions by maximizing expected log-scoring, i.e. $\mathbb{E}_P log(p(X|Z,\vartheta,J)$ where ϑ is the parameter of interest and J represent non-data dependent choices that affect the score. The loss can be further motivated by the fact that making better or more informed choices involves a cost i.e. computing power, that is non negligible. We denote such a cost by $\varepsilon(J)$. Taking quadratic approximation to expected log-loss around the pseudo-true of the parameter of interest and subtracting from the log loss based on $p^{\star}(X|Z,\vartheta)$, we have that:

$$\begin{aligned}
\mathbb{E}_P log\left(\frac{p^{\star}(X|Z,\vartheta_0)}{p(X|Z,\hat{\vartheta},J)}\right) &= \mathbb{E}_P log\left(\frac{p^{\star}(X|Z,\vartheta_0)}{p(X|Z,\vartheta,J)}\right) - \frac{1}{2}(\hat{\vartheta}-\vartheta)^T \frac{\partial^2 log(p(X|Z,\tilde{\vartheta},J))}{\partial\vartheta\vartheta^T}(\hat{\vartheta}-\vartheta) \\
&= \varepsilon(J) - \frac{1}{2}(\hat{\vartheta}-\vartheta)^T \frac{\partial^2 log(p(X|Z,\tilde{\vartheta},J)}{\partial\vartheta\vartheta^T}(\hat{\vartheta}-\vartheta)
\end{aligned}$$

where the first term of the right hand side is the relative loss due to the auxiliary decisions, J, and does not depend on N_0.

It is important to recognize that there is no reason to believe that $\varepsilon(J)$ diminishes as the sample size N_0 grows, since it is non-data revisable. For example, in the context of the approximation of dynamic equilibrium models, (Ackerberg, Geweke, and Hahn 2009) have shown (in Theorem 4) that consistency for ϑ_0 is achieved only when a measure of approximation, Δ_j[3], converges to zero at a faster rate than $\sqrt{N_0}$. However, there is no intuitive reason why the quality of approximation can be tied to the sample size, and therefore the assumption behind the

[3] $\Delta_j \equiv \max\{\sup_{X_t,\vartheta}|\frac{\partial^k}{\partial\vartheta^k}log(p_j(X|\vartheta,J)) - \frac{\partial^k}{\partial\vartheta^k}log(p^{\star}(X|Z,\vartheta))|\}$

asymptotic result become tenuous. What is more is that we assume away the case of loss of identification that arises due to approximations or equilibrium selection. Although this is very important per se, it is not consequential in terms of the log-score itself as different values of ϑ lead to the same $p(X|Z,\vartheta)$.

The above problem can be embedded in a Principal-Agent framework, in which the Principal delegates parameter estimation to the Agent and supplies the sample of data and a prior over the form of the distribution of the data. The latter is provided for two reasons. First, there is a monetary loss incurred for obtaining the otherwise unknown full likelihood. This necessitates the introduction of prior information, usually coming from previous experience of the Principal. In fact, the cost $\varepsilon(J)$ calibrates a set of prior models as follows:

$$\mathcal{B}(p^{\star}(X|\vartheta),\varepsilon(J)) \quad := \quad \{p \in \mathcal{P} : \mathbb{E}log(\frac{p^{\star}(X|Z,\vartheta)}{p}) < \varepsilon(J)\}$$

Any prior p that yields a loss less than $\varepsilon(J)$ is therefore admissible. Second, the Principal seeks to guard against possible model overfit by the agents. One way of achieving this is by data holdouts (see for example (Schorfheide and Wolpin 2016)). Another way is to prefer models that are close as possible to prior information about the conditional density of the observables. We follow the latter approach.

Moreover, the compensation scheme is such that the agent is rewarded based on log score i.e. the "in sample" predictive performance. The Principal gains by a high log score as this implies better choice of policies and higher welfare. On the other hand, she penalizes models that are too far from the prior distribution. This can be represented by a two stage game in which the Principal first chooses the family of densities that satisfy the restrictions coming from economic theory, and then delegates model fit to the econometrician. This framework also fits real world situations in which the government or manager has some more experience in what is a good statistical model that has a good prior predictive performance. On the other hand, the economic model is defined up to a set of equilibrium conditions. The Principal therefore solves the following program:

$$\min_{h(X|Z,\varphi)\in\mathcal{H}_{\theta}} \int h(X|Z,\psi)log(\frac{h(X|Z,\psi)}{f(X|Z,\varphi)})h(Z)d(X,Z) \tag{3.1}$$

where

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

$$b) \qquad f(X|Z,\varphi) \in \mathcal{B}(p^\star(X|\vartheta), \varepsilon(J))$$

$$c) \qquad \psi^\star = \underset{\psi \in \Psi}{argmin} \int \mathbb{P}(X,Z) \log\left(\frac{\mathbb{P}(X,Z)}{h(X,Z,\psi)}\right) d(X,Z)$$

In the information projections literature the minimization problem in 3.1 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.

As already mentioned, the set $\mathcal{H}_\theta$ is the set of admissible densities i.e. the densities that by construction satisfy the moment conditions. Solving the game backwards, they should satisfy these conditions at $\psi^\star$ i.e. taking into account the optimal choice of the econometrician. The second qualifying statement (b) serves as a participation constraint, which by assumption always binds for high enough cost ($\varepsilon(J)$) of actually solving the model and not using prior information.

In stage two, the econometrician therefore solves the standard parameter estimation problem:

$$\max_{\psi \in \Psi} \int \log(h^\star(X,Z|\psi)) d\mathbb{P}(X,Z) \tag{3.2}$$

Note that this game, whose extensive form is in figure 3.1, is not identical to the standard game theoretic setup in the robustness literature i.e. (Hansen and Sargent 2005). In our case the game is not zero sum, and therefore it does not belong to the min-max class of games. Moreover, the game is trivially sequential in the sense that the Principal commits to a choice of conditional density from time 0, while the Econometrician acts at time $N > 0$. Since the Econometrician is not forward looking, there is not reason to keep track of the cost of keeping the Principal's promise, as is typical in dynamic Stackelberg (Ramsey) games, something that greatly simplifies the setup.

The problem of the Principal can be conveniently rewritten such that the choice of density $h(Y|Z,\theta)$ 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

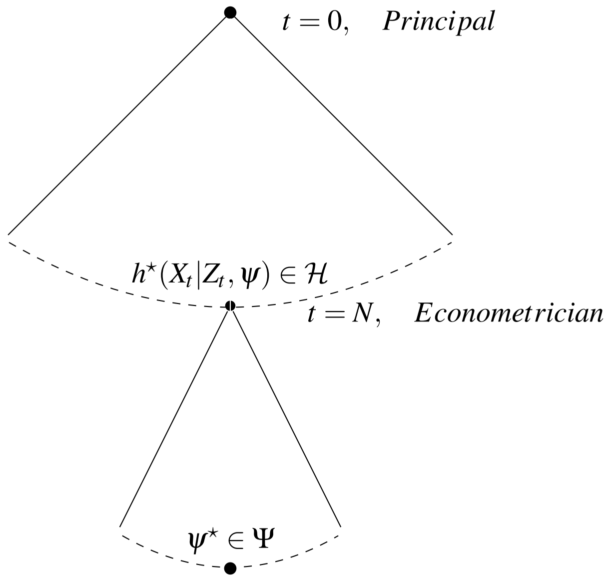


Figure 3.1: Stackelberg Game

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:

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

where

$$a)\qquad \mathbb{M}:=\left\{\mathcal{M}\in\mathcal{L}_p:\mathbb{E}_{f_{(Z,\psi)}}\mathcal{M}(X,Z,\theta)m(X,Z,\theta)=0,\quad \mathbb{E}_{f_{(Z,\psi)}}\mathcal{M}(X,Z,\theta)dY=1\right\}$$

$$b)\qquad f(X|Z,\varphi)\in\mathcal{B}(p^\star(X|\vartheta),\varepsilon(J))$$

In turns out that the optimal choice for the 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:

$$h(X|Z,\psi)=f(X|Z,\varphi)\exp\left(\lambda(Z)+\mu(Z)'m(Y,\vartheta)\right) \tag{3.3}$$

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. This density will be used to perform pseudo-maximum likelihood estimation of the equilibrium model under consideration.

Below, 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 distribution.

3.2.1 An Example

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, that is, 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 and for analytical tractability that they are not correlated. Their joint density conditional on $\mathcal{F}_t$ is therefore:

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

Given the assumption on the utility function, $\mathbb{E}(R_{t+1}c_{t+1}|\mathcal{F}_t) = \frac{c_t}{\beta}$ or $Cov(R_{t+1}c_{t+1}|\mathcal{F}_t) = \frac{c_t}{\beta}(1 - R_t\beta\rho_c\rho_R)$. This is a restriction on the conditional covariation of consumption and interest rates and the new density $h(c_{t+1}, R_{t+1}|\mathcal{F}_t)$ is therefore:

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

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

$$\begin{aligned} \mathcal{M} &= \left[N\left(\begin{pmatrix} \rho_c c_t \\ \rho_R R_t \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ * & 1 \end{pmatrix} \right) \right]^{-1} N\left(\begin{pmatrix} \rho_c c_t \\ \rho_R R_t \end{pmatrix}, \begin{pmatrix} 1 & \frac{c_t}{\beta}(1 - R_t\beta\rho_c\rho_R) \\ * & 1 \end{pmatrix} \right) \\ &= \exp\left(-\frac{1}{2} \left((c_{t+1} - \rho_c c_t, R_{t+1} - \rho_R R_t) \begin{pmatrix} 1 & \frac{c_t}{\beta}(1 - R_t\beta\rho_c\rho_R) \\ * & 1 \end{pmatrix} \begin{pmatrix} c_{t+1} - \rho_c c_t \\ R_{t+1} - \rho_R R_t \end{pmatrix} \right) \right) \end{aligned}$$

In Appendix A, we illustrate how the same expression for $\mathcal{M}$ can be obtained formally using a conditional density projection, that is, solving 3.1 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 this approach. More particularly, in the next section we provide the relevant asymptotic theory which is an extension of the asymptotic theory developed for empirical likelihood (see e.g. (Newey and Smith 2004, Kitamura, Tripathi, and Ahn 2004)). The main difference arise due to the fact that we project on a general possibly misspecified density. Moreover, explicitly acknowledging for estimating the parameters of the density yields some useful insight to the behaviour of the estimator.

3.3 Frequentist Inference

This section illustrates asymptotic results, that is consistency and asymptotic distribution for $\psi \equiv (\vartheta, \psi)$. The properties of the estimator, as expected, depend crucially on the distance between the prior and the true population conditional density. In the case that the distance vanishes for large N_0, we provide an explicit shrinkage formulation in the GEL setting[4].

Before stating the main results, we make certain assumptions that are fairly standard in extremum estimation and are necessary and sufficient for the Propositions to be valid.

For $\{X_i, Z_i\}_{i=1,n\geq 1}^{N_0}$ a stationary ergodic sequence, the following assumptions hold:

Assumptions I

1. **(COMP)** The sets $\Theta \subset \mathbb{R}^k, \varphi \subset \mathbb{R}^l$,$\mathcal{M} \subset \mathbb{R}^M$ are compact, and therefore $\Psi \equiv \Theta \times \varphi \subset \mathbb{R}^{k+l}$ is compact.

2. **(ID)**There exists a unique vector $(\vartheta_0 \in int(\Theta), \varphi_0 \in int(\varphi))$ such that $(\vartheta_0, \varphi_0) = \arg\max_{\Psi} \mathbb{E} \log h(x_t, \vartheta_0, \varphi_0)$

3. **(BD-1**a)$\forall l \in 1..M$ and for $d \leq 4, P \in (F_\varphi, \mathbb{P})$:
$\mathbb{E}_{P|z} \sup_\psi ||m_l(x, \vartheta)||^d, \mathbb{E}_{P|z} \sup_\psi ||m_{l,\vartheta}(x_t, \vartheta)||^d$, and $\mathbb{E}_{P|z} \sup_\psi ||m_{l,\vartheta\vartheta}(x_t, \vartheta)||^d$ are finite, P_z-a.s.

[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).

4. (**BD-1**b)$\sup_{\psi} \mathbb{E}_{\mathbb{P}_{z_i}} ||e^{\mu_i' |m(x,z_i,\vartheta)|}||^{1+\delta} < \infty$ for $\delta > 0$, $\forall \mu_i > 0$ and $\forall z_i$[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 φ. In case φ is infinite dimensional, the conditions have 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. $\Phi_n \subseteq \Phi_{n+1} \ldots \subset \Phi$, 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 the equilibrium conditions.

Recall that the problem for the econometrician is to maximize the likelihood provided by the Principal (3.2). The empirical analogue is therefore the following:

---

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

$$\max_{(\theta,\varphi)\in\Theta\times\varphi} \frac{1}{N_0}\sum_{i=1..N_0} \log\left(f(x_i|z_i,\varphi)\exp(\mu_i' m(x_i,z_i,\vartheta)+\lambda_i\right) \quad (3.4)$$

subject to: (3.5)

$$\int f(X_i|Z_i,\varphi)\exp(\mu_i' m(X_i,Z_i,\vartheta) m_s(X_k,Z_i,\vartheta)dX = 0, \forall l = 1..\dim(m), i = 1..n \quad (3.6)$$

$$\int f(X_i|Z_i,\varphi)\exp(\mu_i' m(X_i,Z_i,\vartheta)dX = 1 \forall i = 1..n \quad (3.7)$$

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. For **M** the Jacobian of the moment conditions, the first order conditions are:

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

where:

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

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 $\hat{\varphi}$ in the neighborhood of $\varphi_0^\star$. 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.

Below we present consistency results for both the case of misspecification and correct specification, and the asymptotic distribution under the former case.

3.3.1 Consistency and Asymptotic Normality

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.

Proposition 5. *Consistency for* $\psi_0^\star$
Under Assumptions I, for N_s,$N_0 \to \infty$ *such that* $\frac{N_0^{\bar{\gamma}+1}}{N_s} \to c$ *with* $c > 0$ *and* $\bar{\gamma} > 1 + \frac{2}{d}$*:*

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

Proof. See the Appendix

□

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

Corollary 3.1. *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 □

Given consistency, we can proceed in deriving the limit 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 6. *Asymptotic Normality*
Under asymptotic correct specification and Assumption I:

$$N_0^{\frac{1}{2}}(\psi - \psi^*) \underset{d}{\to} N(0, I(\psi_0)^{-1})$$

where $I(\psi_0)^{-1}$ *is the semi-parametric lower bound,* $I(\psi_0)^{-1} = \mathbb{E}(G(\psi, z)'\mathbb{V}_g(\psi, z)^{-1}G(\psi, z))$.

Proof. See the Appendix

□

In Appendix B we derive the exact form of the variance covariance matrix of the estimator. Given a finite number of conditional moment restrictions, this is the lowest variance bound that can be achieved, see for example (Chamberlain 1987). With regard to the Jacobian terms,

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

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

$$\mathbb{E}G_{i,\vartheta_I\vartheta'} = \mathbb{E}M_i(\vartheta)'V_m^{-1}(\vartheta)M_i(\vartheta) \tag{3.8}$$

$$\mathbb{E}G_{i,\vartheta\varphi'} = \mathbb{E}_z M_i(\vartheta)V_m^{-1}\mathbb{E}(m_i(\vartheta)\otimes\mathfrak{s}_i(\varphi)'|Z) \tag{3.9}$$

$$= \mathbb{E}_z M_i'(\vartheta)\mathfrak{B}_i(\psi) \tag{3.10}$$

$$\mathbb{E}G_{i,\varphi\varphi'} = \mathbb{E}_z\mathfrak{s}_i(\varphi)\mathfrak{s}_i(\varphi)' \tag{3.11}$$

Notice also 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, 3.11 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:

$$\bar{V}_{11} = \mathbb{E}_z M_i(\vartheta)'V_m^{-1}M_i(\vartheta)$$

$$\bar{V}_{22} = \mathbb{E}_z(\mathfrak{s}_i(\varphi)+\mathfrak{s}_i^P(\varphi))(\mathfrak{s}_i(\varphi)+\mathfrak{s}_i^P(\varphi))'$$

$$\bar{V}_{12} = \mathbb{E}_z M_i(\vartheta)'\mathfrak{B}_i(\psi)$$

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 :

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

Interestingly, the above conditions have an intuitive interpretation. To see this, notice that 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:

$$\bar{V}_0 = \begin{pmatrix} H'(V_{\mathfrak{s}}+V_{\mathcal{U}})^{-1}H & H' + \frac{\partial \mathcal{U}}{\partial \phi} \\ H + \frac{\partial \mathcal{U}}{\partial \phi}' & 4V_{\mathfrak{s}} \end{pmatrix}$$

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

$$\bar{V}_0 = \begin{pmatrix} V'_{\mathfrak{s}}(V_{\mathfrak{s}}+V_{\mathcal{U}})^{-1}V_{\mathfrak{s}} & V'_{\mathfrak{s}} + \frac{\partial \mathcal{U}}{\partial \phi} \\ V_{\mathfrak{s}} + \frac{\partial \mathcal{U}}{\partial \phi}' & 4V_{\mathfrak{s}} \end{pmatrix}$$

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 the the Cramer - Rao bound is attained as $\mathcal{U} = 0$.

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 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.

3.4 Shrinkage Towards the Statistical Model

3.4.1 Parametric Case

Since we have limited the scope of our analysis to the case of finite dimensional φ, it is instructive to see what happens under misspecification. For simplicity 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 deonote $m_P \equiv \int m(X,Z) dP(X,Z)$ 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 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:

$$0 \quad = \quad (M_{\mathbb{P}_n} - M_{H_n})' V_{\kappa,f_n}^{-1} N_0^{\frac{1}{2}} (m_{f_n} - m_{\mathbb{P},n}) + (M_{\mathbb{P}_n} - M_{H_n})' V_{\kappa,f_n}^{-1} N_0^{\frac{1}{2}} m_{\mathbb{P}_n} + \ldots \qquad (3.12)$$

$$\ldots \quad + (M_{f_n}' V_{f_n}^{-1} - M_{\mathbb{P}_n}' V_{\mathbb{P}_n}^{-1}) N_0^{\frac{1}{2}} m_{\mathbb{P}_n} + M_{\mathbb{P}_n}' V_{\mathbb{P}_n}^{-1} N_0^{\frac{1}{2}} m_{\mathbb{P}_n} \qquad (3.13)$$

The first three terms are functions of the distance between the proposed and the true $f(x|z)$. Furthermore, we make use of 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 3. *Influence function for plug-in estimator (Wasserman 2006)*
For a general function $W(x,z)$, conditional density $Q(x|z)$ and $\mathcal{L}(x,z) \equiv W(x,z) - \int W(x,z)d\mathbb{P}_z(x|z)$

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

Given Lemma 3, we can the 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 7. *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:

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

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 Jacobians. Given Proposition 1, we can now determine whether the first order estimating equations for ϑ are affected by the misspecification.

Proposition 8. *Local misspecification has first order effects on $\hat{\vartheta}$ only through $\hat{\phi}$.*
Given Proposition 1, the system of equations in (3.12) 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 9. *Non Regularity and Superefficiency*
For $R(\vartheta) \equiv \frac{\partial}{\partial\vartheta} r(\vartheta)$, $G^{-1} \equiv \begin{pmatrix} G^{11} & G^{12} \\ G^{21} & G^{22} \end{pmatrix}$, $S_1 \equiv [I_{n_1}, 0_{n_1,n_2}]$, $S_2 \equiv [0_{n_2,n_1}, I_{n_2}]$,

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

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

   where $\mathcal{Z}_r \equiv S_1\mathcal{Z} - G^{12}(\psi_0)R(\vartheta_0)(R(\vartheta_0)'G^{22}(\psi_0)R(\vartheta_0))^{-1}R(\vartheta_0)'(S_2(\mathcal{Z}+h))$

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 semiparametric 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 conventional semiparametric optimal feasible GMM.

3.4.2 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.

3.5 Discussion and Simulation Evidence

3.5.1 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. The type of misspecification that is most likely to arise is the type of the distribution. Again, there might be ample of previous evidence on how skewed to fat tailed the distribution is. For example, we know that financial data have heavy tails, and it would be unwise for any practitioner not to account for that. Below, we provide evidence of how severe the effects on MSE can be in a simple setting.

3.5.2 Monte Carlo experiments and MSE

In order to facilitate the Monte Carlo exercise, we consider a fairly general DGP which at the same time satisfies a non linear moment condition. Furthermore, to ease computation we utilize the representation of first order conditions in 3.12[7]. In the Appendix, we also discuss the

[7] 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 $\int |M(x,\vartheta)||p(x) - h(x,\varphi,\vartheta)|dx$.

computational details and provide the algorithm to implement the estimation. To further reduce the computational burden, in the Monte Carlo experiment we focus on the special case of an unconditional moment restriction. In the application to a typical equilibrium model we allow for conditional moment restrictions.

3.5.2.1 Design of 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 3.1: 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	$N(0,4)$,$U(0,1)$	$N(0,4)$,$U(0,1)$

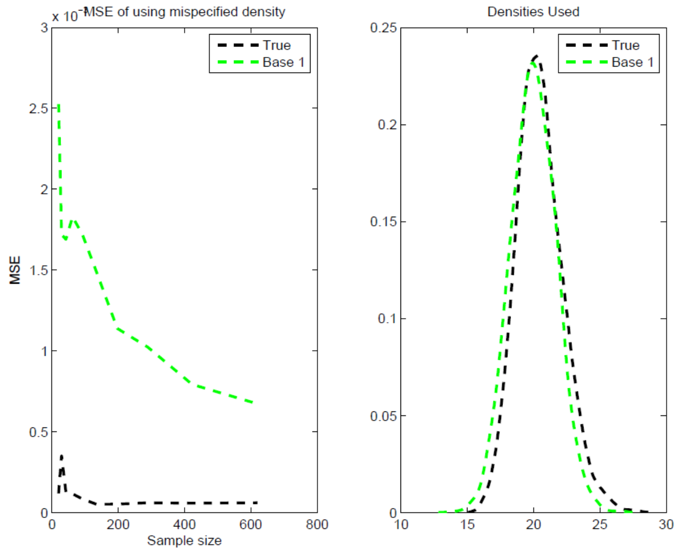


Figure 3.2: Monte Carlo Case 1

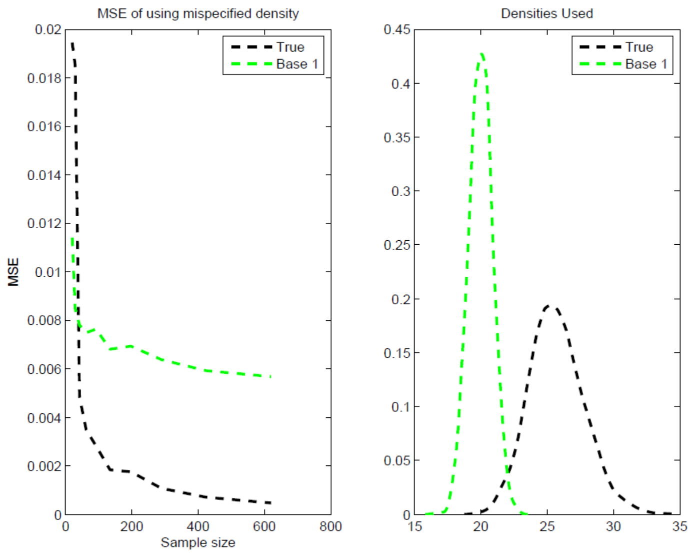


Figure 3.3: Monte Carlo Case 2

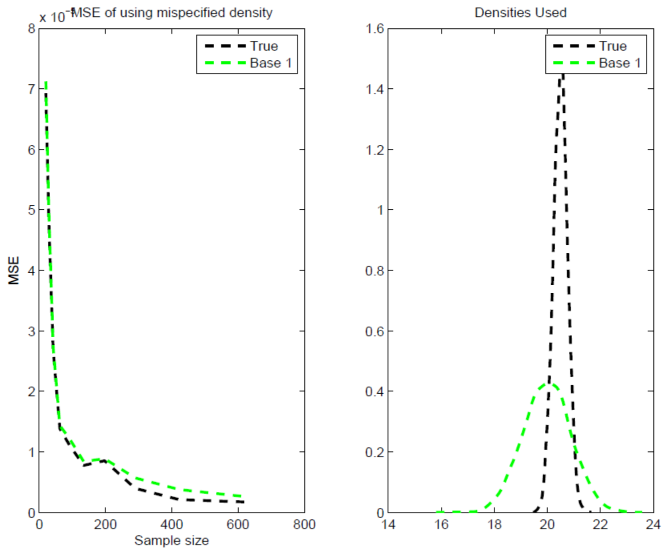


Figure 3.4: Monte Carlo Cases 3

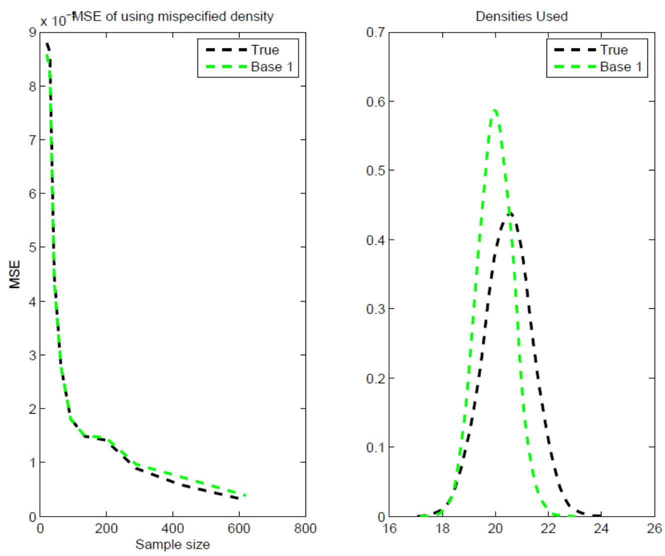


Figure 3.5: Monte Carlo Case 4

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.

What we also need to mention is the fact that we do not estimate the parameters of the proposed density. We nevertheless predict that allowing for estimation of the parameters of the density will have three implications. First, there will be additional noise coming from the estimation of the nuisance parameters. Second, the proposed density will be as close as possible to the true density. Third, we should notice the shrinkage properties characterized in Proposition 9. Moreover, given the fact that the densities used are not locally misspecified, the first order conditions for ϑ are no longer unbiased, in the sense of Proposition 8. The bias is therefore larger in this case.

3.5.3 Application to 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:

$$\begin{aligned}
Y_t &= A_t K_t^{\theta} H_t^{1-\theta} && (3.14)\\
K_{t+1} &= Y_t - C_t && (3.15)\\
\gamma C_t H_t &= (1-\theta) Y_t I_t && (3.16)\\
\frac{1}{C_t} &= \beta \mathbb{E}_t \left\{ \frac{1}{C_{t+1}} \left(\theta \left(\frac{Y_{t+1}}{K_{t+1}} \right) \right) \right\} && (3.17)\\
log(I_{t+1}) &= \rho_I log(I_t) + logN(0, \sigma_I^2) && (3.18)\\
log(A_{t+1}) &= \rho_I log(A_t) + logN(0, \sigma_A^2) && (3.19)\\
& && (3.20)
\end{aligned}$$

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 [8].

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 $\underset{n_\psi \times 1}{\phi}$. In the Appendix 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:

$$\frac{1}{C_t} = \beta\theta\mathbb{E}_t \left\{ \frac{1}{C_{t+1}} A_{t+1} \left(\frac{H_{t+1}}{K_{t+1}} \right)^{1-\theta} \right\} \tag{3.21}$$

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 :

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

Table 3.2: 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.

Although the stochastic growth model we have used is quite basic, we should note that te approach suggested by this paper can be extended to more elaborate equilibrium models, as long as moment conditions can be constructed trough simulation. In Appendix A we also discuss the potential application to models with non-differentiabilities, which imply that first order conditions cannot be readily employed. We pursue the discussion in the context of discrete choice modeling.

3.6 Conclusion and Future Research

In this paper we have proposed an alternative approach to estimating a probability model that satisfies (un)conditional moment restrictions coming from equilibrium models. The motivation comes from the fact that solving these models for the equilibrium 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 and finite sample 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. In is worthwhile to note that the results of this paper are general and are 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 some simulation evidence for the performance of this method at various sample sizes under non local misspecification. A comparison of the performance of this method to full information methods is left for future work.

3.7 Appendix A

3.7.1 Analytical derivations for Example 1

The well defined optimization problem is the following:

$$\min_{h \in H} \int \log(\frac{h(x|z,\psi)}{f(x|z,\varphi)}) h(x|z,\psi) dx \quad s.t \quad \int h(x|z,\psi) m(x,\vartheta) dx = 0$$
$$\int h(x|z,\psi) dx = 1$$

The Lagrangian is

$$\min_{h \in H} \int \log(\frac{h(x|z,\psi)}{f(x|z,\varphi)}) h(x|z,\psi) dx + \mu \int h(x|z,\psi) m(x,\vartheta) dx + \lambda (\int h(x|z,\psi) dx - 1)$$

The solution to this variational problem is the following, $h(x|z,\psi) = f(x|z,\varphi)\exp(\mu' m(x,\vartheta) + \lambda)$ where $\mu = \arg\min \int f(x|z,\varphi)\exp(\mu' m(x,\vartheta))$ and $\lambda = -\log(\int f(x|z,\varphi)\exp(\mu' m(x,\vartheta)))$.

In the particular example given in the paper, μ is the solution to $\frac{\mu}{1-\mu^2} = \frac{c_t}{\beta}(1 - R_t \rho_c \rho_R \beta)$. To see this, notice that, suppressing λ, the perturbation, $\exp(\mu' m(x,\vartheta) + \lambda)$ is proportional to
$$\exp\left(-\frac{1}{2}\left(\begin{pmatrix} c_{t+1} - \rho_c c_t \\ R_{t+1} - \rho_R R_t \end{pmatrix}' \begin{pmatrix} 0 & -\mu_t \\ -\mu_t & 0 \end{pmatrix} \begin{pmatrix} c_{t+1} - \rho_c c_t \\ R_{t+1} - \rho_R R_t \end{pmatrix}\right) - 2\mu_t \frac{c_t}{\beta}(1 - R_t \rho_c \rho_R \beta)\right)$$
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 $\begin{pmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \end{pmatrix} \equiv \begin{pmatrix} c_{t+1} - \rho_c c_t \\ R_{t+1} - \rho_R R_t \end{pmatrix}$ the problem becomes

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

We therefore have that the F.O.C

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

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

Then, for the Normal scaling constant C,

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

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

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

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

3.7.2.1 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 $F(\mu) = \int f(x,\varphi)\exp(\sum_j m_j(x,\vartheta)) \approx \frac{1}{n_s}\sum_{i=..N_s}\exp(\sum_j m_j(x_i,\vartheta))$. 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.

3.7.3 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. 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, $c_t = -\frac{1}{\sigma} r_t$.:

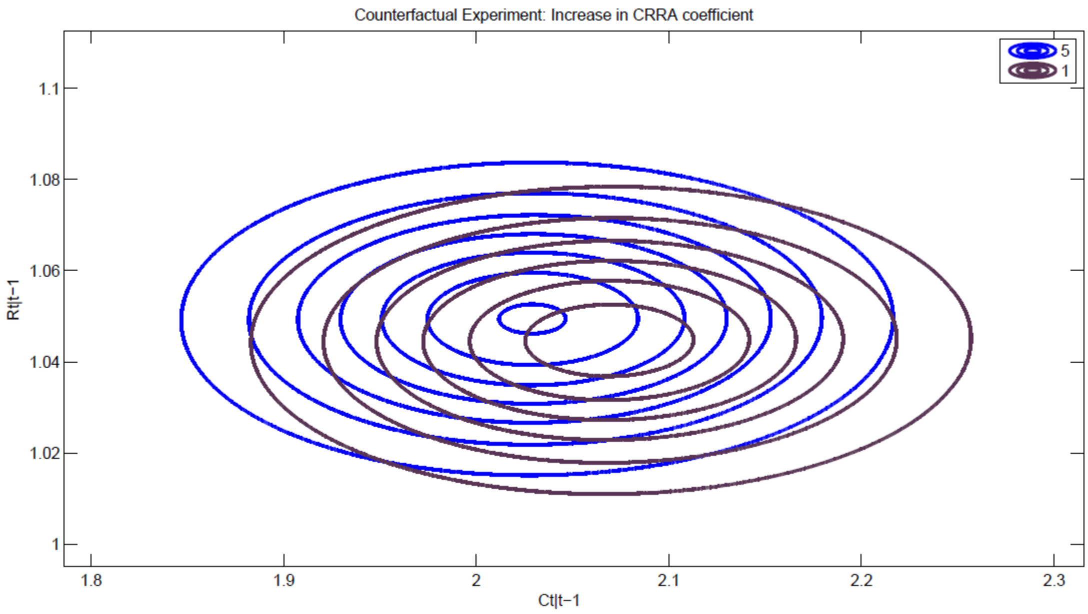


Figure 3.6: Increase in Risk Aversion Coefficient

3.7.4 Reduced form coefficients in stochastic growth model

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

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

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

3.7.5 A Note On Non Differentiability

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. As already mentioned, the approach of (Su and Judd 2012) is instructive on how we can cast a discrete choice problem in our framework. 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). For now, let me briefly introduce the MPEC approach here, and how it can be casted in our framework. Adapting to the notation of (Su and Judd 2012), let me define the equation $b(\vartheta, \sigma) = 0$ to represent equilibrium conditions, where σ are the policy functions. Then the set $\Sigma(\vartheta) : \{\sigma : b(\vartheta, \sigma) = 0\}$ is the set of policy functions which are consistent with the equilibrium defining equation, and $\hat{\sigma}(\vartheta)$ is an element of this set. In the case of unique equilibrium, this is the unique policy function consistent with equilibrium and for the sake of brevity let me focus on this case. The maximum log likelihood problem is equal to

$$\begin{aligned} \hat{\vartheta} &= \arg\max_{\vartheta \in \Theta} \frac{1}{n} L(\vartheta, \hat{\sigma}(\vartheta); X) \\ &= \arg \max_{\hat{\sigma}(\vartheta) \in \Sigma(\theta)} \frac{1}{n} L(\vartheta, \hat{\sigma}(\vartheta); X) \end{aligned}$$

This can be a difficult problem in terms of computation, as the policy function has to be computed at each iteration. The MPEC approach deals with a mathematically equivalent, but possibly easier problem, that of:

$$\begin{aligned} \hat{\vartheta} &= \arg\max_{\vartheta, \sigma} \frac{1}{n} L(\vartheta, \sigma; X) \\ s.t. \quad & b(\vartheta, \sigma) = 0 \end{aligned}$$

Clearly, in this problem, the policy function is not solved for ϑin the likelihood function. Moving to the discrete choice problem, e.g. (Rust 1987), denote the discrete choice sequence by $\{d_t, d_{t+1}, d_{t+2}...\}$, 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, it is the solution 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, that is the Markov transition density is factorized in the following manner,

$$p(x',\varepsilon'|x,\varepsilon,d,\vartheta) = p_2(\varepsilon'|x',\vartheta_2)p_3(x'|x,d,\vartheta_3)$$

we notice that

$$\begin{aligned} \mathbb{E}V(x) &= \int_\varepsilon V(x,\varepsilon)p_2(\varepsilon|x,\vartheta_2)d\varepsilon) \\ &\mathit{and} \\ \mathbb{E}V(x,d) &= v(x,d,\theta_1) + \beta \int_{x'} \mathbb{E}V(x')p_3(x'|x,d,\vartheta_3)dx') \end{aligned}$$

Under conditional independence, we can write down the log likelihood of a data point $\{X_i\}$as

$$\log l_i(X_{d,i},\vartheta) \quad = \quad \log P(d_i|x_i;\vartheta)p_3(x_i|x_{i-1},d_{i-1},\vartheta_3) \tag{3.22}$$

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. The point that I would like to make here is that looking at 3.22, we can obtain the likelihood in the following sense: 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. $\mathbb{E}(V - T(\mathbb{E}V)) = Em_d(x_{d,i},\vartheta_d)$. The difference, as already pointed out, is that this restriction will be satisfied for the true parameter vector that belongs to this discrete choice problem, and will also a function of φ, i.e. $\vartheta^d = \tau(\varphi)$. More particularly, the tilted density will be of the form:

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

where $\delta(.)$ is the Dirac delta function, used to represent the probability density function of a discrete variable.

It is important to see that the MPEC method, which I have embedded above in our framework, avoids 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.

3.8 Appendix B

Proof. of Proposition 5 :
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. From Proposition 2, plugging in $F_n(x,\varphi,z_i)$, and the definition of$\mu=\arg\inf Q(x,z_i,\vartheta,\phi,\mu)$ for all ϑ,ϕ , $\underset{M\times 1}{\mu}(\phi,\vartheta)$ exists and is unique . Fix $Z=z_i$,, $\forall\delta>0$, 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 (4), 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 (3)

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

and by the Central Limit Theorem for Martingale Difference sequences, 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, $\psi \equiv (\vartheta, \phi) = \arg\sup G(x, \vartheta, \mu)$ where

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

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

The following section seeks to establish uniform convergence results for the objective function. 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) together with (ii) **PC** (Pointwise consistency) and (iii) **SE** (Stochastic Equicontinuity). Regarding (iii), we can proceed as follows:

Regarding (i). since in this section we are dealing with a finite dimensional φ, we assumption 1 **(COMP)** implies total boundedness. Looking at pointwise convergence (ii), consider that under identically distributed data,

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

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

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

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

Therefore the condition that needs to be shown is that,

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

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

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

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

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

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

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

$$\begin{array}{rcl}
\mathbb{E}\frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})| & \leq & \mathbb{E}_z\mathbb{E}_{x|z}\frac{1}{n}\sum_i |\exp(q_i(\bar{\psi}))\nabla q_i(\bar{\psi})| \\
& \leq & \mathbb{E}|\exp(q_i(\bar{\psi}))||\nabla q_i(\bar{\psi})| \\
\text{Cauchy Schwarz} & \leq & \left(\mathbb{E}(|\exp(q_i(\bar{\psi}))|)^2\right)^{\frac{1}{2}} \left(\mathbb{E}(|\nabla q_i(\bar{\psi}))|)^2\right)^{\frac{1}{2}} \\
& < & O_p(1) \times \left(\mathbb{E}2\log f_i(\varphi) + (\mathbb{E}\hat{\mu}_i(\psi)'m_i(\vartheta))^2 + \mathbb{E}\hat{\lambda}_i(\psi)^2\right)^{\frac{1}{2}} \\
\text{BD-1a,BD-2,CS,Prop 6 and Lem 10} & \leq & \left(2\mathbb{E}\log f_i(\varphi) + \mathbb{E}(\hat{\mu}_i(\psi)'\hat{\mu}_i(\psi)')\mathbb{E}m_i(\vartheta)'m_i(\vartheta) + \mathbb{E}\hat{\lambda}_{i,\psi}^2\right)^{\frac{1}{2}} \\
& = & O_p(1)
\end{array}$$

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 **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 (), $\lambda(Z_i) = \mu(Z_i) = 0 \forall i$ and therefore $h(X|Z,\psi) = f(X|Z,\varphi)$. By construction, $\mathbb{E}_H m(X,Z,\vartheta) = \int \mathbb{P}(X,Z)m(X,Z,\vartheta_0^\star)d(X,Z) = 0$. 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 Proposition6: We have the following first order conditions characterizing the estimator:

$$\begin{array}{ll}
\vartheta : & \frac{1}{n}\sum_i \left(\mu_i' M_i(\vartheta) + \mu_{\theta,i}' m_i(\vartheta) + \lambda_{i,\vartheta}\right) = 0 \\
\phi : & \frac{1}{n}\sum_i \left(\frac{s(x_i,z_i,\varphi)}{f(x_i,z_i,\varphi)} + \mu_{\varphi,i}' m_i(\vartheta) + \lambda_{i,\phi}\right) = 0
\end{array}$$

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 $\kappa_{j,i} \underset{p}{\to} -1$.

For $V_{f,m} = \frac{1}{N_s}\sum_j^{N_s} m_j(\vartheta)m_j(\vartheta)'$, let $V_{f,m}^{-1} \equiv (v_1', v_s'..v_{n_m}')'$ and $v_1 \equiv (v_{11}, v_{12}...v_{1n_m})'$.The derivatives terms are the following:

$$
\begin{aligned}
\mu_{i,\vartheta} &= (-\frac{1}{n_s}\sum_j e_j m_j m_j')^{-1}(\frac{1}{n_s}\sum_j z_j M_j)\\
\lambda_{i,\vartheta} &= -\mu_i'\sum_j \tilde{e}_j M_j - \sum_j \tilde{e}_j m_j' \mu_\vartheta\\
\mu_{i,\phi} &= (\sum_j e_j m_j m_j')^{-1}\sum_j e_j m_j \otimes s_j\\
\lambda_{i,\phi} &= -\sum_j \tilde{e}_j \mathfrak{s}_j - \sum_j \tilde{e}_j m_j' \mu_\phi\\
\mu_{i,\vartheta_l\vartheta'} &= \sum_p^{n_m}(v_{l,p,\vartheta'}\frac{1}{N_s}\sum_j^{N_s} z_j M_{p,l,j} + v_{l,p}\frac{1}{N_s}\sum_j^{N_s}(z_j D_{l,p,j} + z_{j,\vartheta'} M_{p,l,j}))\\
&= \sum_p^{n_m}(\frac{1}{N_s}\sum_j^{N_s}(v_{l,p,\vartheta'} z_j + v_{l,p} z_{j,\vartheta'})M_{p,l,j} + \frac{1}{N_s}\sum_j^{N_s} v_{l,p} z_j D_{l,p,j})\\
\lambda_{i,\vartheta_l\vartheta'} &= -\frac{2}{n_s}\sum_j \tilde{e}_j \mu_\vartheta' M_j - \frac{1}{n_s}\sum_j \tilde{e}_j D_{l,\vartheta',j}\ldots\\
\ldots &\quad -\sum_j \tilde{e}_j (m_j'\mu_\vartheta + \mu'(M_j) - \sum_j \tilde{e}_j M_j)' m_j' \mu_\vartheta - \sum_j \tilde{e}_j m_j' \mu_{\vartheta_l,\vartheta'}\\
\mu_{i,\phi_l\phi'} &= [\sum_j \kappa_j m_j m_j']^{-2}[(\frac{1}{n_s}\sum_j m_j \otimes \mathfrak{s}_{l,\phi'} - \mu_\phi \frac{1}{n_s}\sum_j \kappa_j m_j m_j' \otimes \mathfrak{s}_{l,j} - \ldots\\
\ldots &\quad \mu \frac{1}{n_s}\sum_j \kappa_j m_j m_j' \otimes \mathfrak{s}_{l,j,\varphi})\sum_j \kappa_j m_j m_j' \ldots\\
\ldots &\quad -(\frac{1}{n_s}\sum_j m_j \otimes \mathfrak{s}_{l,j} - \mu \frac{1}{n_s}\sum_j \kappa_j m_j m_j' \otimes \mathfrak{s}_{l,\phi',j})\frac{1}{n_s}\sum_j \kappa_j m_j m_j' \otimes \mathfrak{s}_{l,j}]\\
\lambda_{i,\phi_l\phi'} &= -\frac{1}{n_s}\sum_j \tilde{e}_j \mathfrak{s}_{l,\phi',j} + (\frac{1}{n_s}\sum_j \tilde{e}_j \mathfrak{s}_l)'(\frac{1}{n_s}\sum_j \tilde{e}_j \mathfrak{s}_l)\\
\lambda_{i,\vartheta_l\phi'} &= -\sum_j e_j' M_{l,j}' \mu_\phi - \mu'\sum_j e_j M_{l,j} \otimes \mathfrak{s}_j + \mu'\sum_j e_j \mathfrak{s}_j \sum_j \tilde{e}_j M_l, j - \sum_j \tilde{e}_j m_j \otimes \mathfrak{s}_j \mu_{\vartheta_l} - \sum_j \tilde{e}_j m_j \mu_{\vartheta_l\phi}\\
\mu_{i,\vartheta\phi} &= (\frac{1}{n_s}\sum_j e_j m_j m_j')^{-1}\frac{1}{n_s}\sum_j z_j M_j \frac{1}{n_s}\sum_j e_j m_j m_j' \otimes \mathfrak{s}_j (\frac{1}{n_s}\sum_j e_j m_j m_j')^{-1}\ldots\\
\ldots &\quad -(\frac{1}{n_s}\sum_j e_j m_j m_j')^{-1}\frac{1}{n_s}\sum_j z_j M_j \otimes \mathfrak{s}_j
\end{aligned}
$$

Under correct specification, the unconditional moments of the above derivatives are as follows:

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}_n}\mu_{i,\vartheta} &\xrightarrow{p} -V_{P,m}^{-1} M_P\\
\mathbb{E}_{\mathbb{P}_n}\lambda_{i,\vartheta} &\xrightarrow{p} 0\\
\mathbb{E}_{\mathbb{P}_n}\mu_{i,\phi} &\xrightarrow{p} Projc(\mathfrak{s}|m)\\
\mathbb{E}_{\mathbb{P}_n}\lambda_{i,\phi} &\xrightarrow{p} 0\\
\mathbb{E}_{\mathbb{P}_n}\mu_{i,\vartheta_l\vartheta'} &\xrightarrow{p} \sum_p^{n_m}(v_{l,p,\vartheta'} M_{p,l,j} + v_{l,p} D_{l,p,j})\\
\mathbb{E}_{\mathbb{P}_n}\lambda_{i,\vartheta\vartheta} &\xrightarrow{p} -2\mu_\vartheta' M_l - D_{l,\vartheta}
\end{aligned}
$$

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_n}\mu_{i,\phi_l\phi'} &\xrightarrow{p} \frac{\partial}{\partial\phi'}Projc(\mathfrak{s}_l|\mu)\\
\mathbb{E}_{\mathbb{P}_n}\lambda_{i,\phi_l\phi'} &\xrightarrow{p} 0\\
\mathbb{E}_{\mathbb{P}_n}\lambda_{i,\vartheta_l\phi'} &= -M_l'Projc(\mathfrak{s}|m) - Projc(\mathfrak{s}|m)'M_l\\
\mathbb{E}_{\mathbb{P}_n}\mu_{i,\vartheta_l\phi'} &\xrightarrow{p} V_m^{-1}M_l\mathbb{E}(mm'\otimes\mathfrak{s}')V_m^{-1} - V_m^{-1}\mathbb{E}M_l\otimes\mathfrak{s}'
\end{aligned}$$

where $Projc := V_m^{-1}\mathbb{C}ov(\mathfrak{s},m)$.

The estimator has this final implicit form:

$$\underset{p\times M}{G_n} = 0$$

$$\underset{p\times M}{G_n} = \begin{bmatrix}
\frac{1}{n}\sum_i \underbrace{\left(\frac{1}{n_s}\sum_j \kappa_{j,i}m_{j,i}(\vartheta)m_{j,i}(\vartheta)'\right)}_{A_1}^{-1} \left(\underbrace{M_i(\vartheta) - \frac{1}{n_s}\sum_j \tilde{e}_{j,i}M_{j,i}(\vartheta)}_{B_1}\right)' \left(\underbrace{\frac{1}{n_s}\sum_j m_{j,i}(\vartheta)}_{C_1}\right)\\
\frac{1}{n}\sum_i \left(\underbrace{\frac{1}{n_s}\sum_j e_{j,i}m_{j.i}(\vartheta)m_{j,i}(\vartheta)'}_{A_2}\right)^{-1} \left(\underbrace{\frac{1}{n_s}\sum_j e_{j,i}m_{j,i}(\vartheta)\otimes s_{j,i}(\psi)}_{B_{22}}\right) \left(\underbrace{m_i(\vartheta)}_{C_2}\right)
\end{bmatrix} + ..$$

$$... + \begin{bmatrix}
\frac{1}{n}\sum_i \left(\frac{1}{n_s}\sum_j e_{j,i}m_{j.i}(\vartheta)m_{j,i}(\vartheta)'\right)^{-1} \left(\underbrace{\frac{1}{n_s}\sum_j z_{j,i}M_{j,i}(\vartheta)}_{B_{21}}\right) (m_i(\vartheta))\\
\frac{1}{n}\sum_i \underbrace{(\mathfrak{s}(x_i,z_i,\varphi) - \frac{1}{n_s}\sum_j \tilde{e}_{j,i}\mathfrak{s}(x_j,z_i,\varphi))}_{B_{23}}
\end{bmatrix}$$

The way forward would be to decompose $n^{\frac{1}{2}}G_{1,i,n} = n^{\frac{1}{2}}A_{i,0} + n^{\frac{1}{2}}\Delta_i$ and show that $n^{\frac{1}{2}}\Delta_i = o_p(1)$. To start with, we investigate the convergence of individual quantities.

$$\hat{A}_1 = \frac{1}{n_s}\sum_j \kappa_{j,i}m_{j,i}(\vartheta)m_{j,i}(\vartheta)' \xrightarrow{p} A_1$$

by *iid* sampling and domination assumptions $\mathbf{BD-1}$. Regarding A_2 and C_1, they converge to $\mathbb{E}_h m(\vartheta)m(\vartheta)' \equiv V_e$ and $C_1 = \frac{1}{n_s}\sum_j m_{j,i}(\vartheta) \xrightarrow{p} \mathbb{E}_{f(x|z_i)}m(\vartheta) = m_F$ respectively. For B_{12}, the fact that we divide by the sum of the weights in each case, makes it convenient, as given the iid assumption, we have that :

$$B_{12} = \frac{1}{n_s}\sum_j \tilde{e}_{j,i}M_{j,i}(\vartheta)$$

$$\overset{p}{\to} \quad \mathbb{E}_{h(x|z_i)}M(\vartheta) \quad = M_H$$

Moreover, using Lemma 3,

$$B_{21} = \frac{1}{n_s}\sum_j z_{j,i}M_{j,i}(\vartheta) \quad \overset{p}{\to} \mathbb{E}_{f(\phi|z_i)}M(\vartheta)) = M_F$$

$$\frac{1}{n_s}\sum_j \tilde{e}_{j,i}\mathfrak{s}(x_j,z_i,\varphi) \quad \overset{p}{\to} \quad \mathbb{E}_{H(\phi|z_i)}\mathfrak{s}(\varphi)) = \mathfrak{s}_H$$

The convergence of the above quantities to well defined random variables will enable us to focus on the desired quantities. 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. We rewrite the above first order conditions in the following form:

$$G_n = \begin{bmatrix} \frac{1}{n}\sum_i \left(\left(\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}\right)\hat{B}_{1,i}\hat{C}_1 + A_{i,1}^{-1}M_i(\vartheta)C_{1,i}\right) + A_{i,1}^{-1}\hat{B}_{11,i}\left(\hat{C}_{i,1} - C_{i,1}\right)\ldots \\ \ldots + \frac{1}{n}\sum_i \left(\ldots + A_{i,12}^{-1}(\hat{B}_{21,i} - B_{2,i})\hat{C}_{2,i}\right) \\ \frac{1}{n}\sum_i \left(\left(\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}\right)\hat{B}_{22,i}\hat{C}_{2,i} + A_{i,2}^{-1}(\hat{B}_{22,i} - B_{2,i})\hat{C}_{2,i} + A_{i,2}^{-1}B_{22,i}\hat{C}_{2,i}\right) \end{bmatrix} + \ldots$$

$$\ldots + \begin{bmatrix} \frac{1}{n}\sum_i \left(\left(\hat{A}_{i,12}^{-1} - A_{i,2}^{-1}\right)\hat{B}_{21,i}\hat{C}_{2,i} + A_{i,1}^{-1}\left(\hat{B}_{i,12}^{-1} - B_{i,12}^{-1}\right)\hat{C}_1\right) + \ldots \\ + \ldots \frac{1}{n}\sum_i \left(A_{i,1}^{-1}B_{i,12}\left(\hat{C}_{i,1} - C_{i,1}\right) + A_{i,1}^{-1}B_{i,12}C_{i,1} + A_{i,2}^{-1}B_{21,i}C_{2,i}\right) \\ + \frac{1}{n}\sum_i \mathfrak{s}(x_i,z_i,\varphi) - \frac{1}{n}\sum_i \frac{1}{n_s}\sum_j \tilde{e}_{j,i}\mathfrak{s}(x_j,z_i,\varphi) \end{bmatrix}$$

Regrouping into terms that vanish (Ξ_1) and terms that do not, (Ξ_2)

$$\Xi_1 = \begin{bmatrix} \frac{1}{n}\sum_i \left(\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}\right)\hat{B}_{1,i}\hat{C}_1 + \frac{1}{n}\sum_i A_{i,1}^{-1}\hat{B}_{11,i}\left(\hat{C}_{i,1} - C_{i,1}\right) + \frac{1}{n}\sum_i A_{i,2}^{-1}(\hat{B}_{21,i} - B_{2,i})\hat{C}_{2,i} \\ + \frac{1}{n}\sum_i \left(\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}\right)\hat{B}_{2,i}\hat{C}_{2,i} + \frac{1}{n}\sum_i A_{i,2}^{-1}(\hat{B}_{2,i} - B_{2,i})\hat{C}_{2,i} \end{bmatrix} + \ldots$$

$$\ldots + \begin{bmatrix} \frac{1}{n}\sum_i \left(\hat{A}_{i,2}^{-1} - A_{i,2}^{-1}\right)\hat{B}_{21,i}\hat{C}_{2,i} + \frac{1}{n}\sum_i A_{i,1}^{-1}\hat{B}_{1,i}\left(\hat{B}_{i,12}^{-1} - B_{i,12}^{-1}\right)\hat{C}_1 + \frac{1}{n}\sum_i A_{i,1}^{-1}B_{i,12}\left(\hat{C}_{i,1} - C_{i,1}\right) \\ - \frac{1}{n}\sum_i \left(\frac{1}{n_s}\sum_j \tilde{e}_{j,i}\mathfrak{s}(x_j,z_i,\varphi) - \mathfrak{s}_{H,i}\right) \end{bmatrix}$$

$$\overset{p}{\to} 0$$

where Lemmas 7 and 11 are systematically applied. For the sake of illustration,consider the first term:

$$||\frac{1}{n}\sum_i \left(\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}\right)\hat{B}_{1,i}\hat{C}_{1,i}|| \quad \leq \quad \max_i ||\hat{A}_{i,1}^{-1} - A_{i,1}^{-1}|| \max_i ||\hat{B}_{1,i}|| ||\frac{1}{n}\sum_i \hat{C}_{1,i}||$$

$$= O_p(\bar{n}_s^{-\frac{1}{2}}) \times O_p(1) \times O_p(n^{\frac{1}{\zeta}}) = O_p(n_z^{-\frac{1}{2}\bar{\gamma}+\frac{1}{\zeta}})$$
$$= o_p(1) \text{ provided that } \bar{\gamma} > \frac{2}{\zeta}$$

where $\mathbf{BD-1a}$and $\mathbf{PD-1}$ are used and for the last term

$$\begin{aligned} ||\frac{1}{n}\sum_i A_{i,2}^{-1} B_{21,i}(\hat{C}_{2,i} - C_{2,i})|| &\leq \max_i ||A_{i,2}^{-1}|| \max_i ||B_{21,i}|| || \frac{1}{n}\sum_i m_i(\vartheta) - \mathbb{E}m_i(\vartheta))|| \\ &= O_p(1) o_p(1)| \\ &= o_p(1) \end{aligned}$$

With regard to non vanishing terms:

$$\hat{\Xi}_2 = \left[\begin{array}{c} \frac{1}{n}\sum_i A_{i,1}^{-1} M_i'(\vartheta) C_{1,i} - \frac{1}{n}\sum_i A_{i,1}^{-1} B_{i,12} C_{i,1} + \frac{1}{n}\sum_i A_{i,2}^{-1} B_{21,i} C_{2,i} \\ +\frac{1}{n}\sum_i(\mathfrak{s}(x_i,z_i,\varphi) - \mathfrak{s}_{H,i}) + \frac{1}{n}\sum_i A_{i,2}^{-1} B_{22,i} C_{2,i} \end{array} \right]$$

Note that under asymptotic correct specification,

$$\begin{aligned} \hat{\Xi}_{23} &= \frac{1}{n}\sum_i A_{i,1}^{-1} M'_{i(\vartheta)} C_{1,i} - \frac{1}{n}\sum_i A_{i,1}^{-1} B_{i,12} C_{i,1}) \\ &= \frac{1}{n}\sum_i A_{i,1}^{-1}(M_{i(\vartheta)} - B'_{i,12})' C_{1,i} \\ &= O(\kappa_n^{-1}) \end{aligned}$$

Similarly, $\hat{\Xi}_{24} = \frac{1}{n}\sum_i \xi_2'(\mathfrak{s}(x_i,z_i,\varphi) - \mathfrak{s}_{H,i}) = O(\kappa_n^{-1})$.

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:

$$\begin{aligned} n^{\frac{1}{2}}\xi'_{p\times 1}\Xi_2 &= n^{-\frac{1}{2}}\sum_i \xi_1' B_{21,i} A_{i,2}^{-1} C_{2,i} + +n^{-\frac{1}{2}}\sum_i \xi_2' B_{22,i} A_{i,2}^{-1} C_{2,i} + O(n^{\frac{1}{2}}\kappa_n^{-1}) \\ &= \hat{\Xi}_{21} + \hat{\Xi}_{22} + o(1) \end{aligned}$$

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 for Martingale Difference Sequences (Billingsley 1961) we conclude. We do not need to actually compute the covariances of the above terms as we can further bound them by their variances using C-S inequality.

With regard to $\Xi_{22}, \mathbb{V}_z(\xi' B_{22} A_2^{-1} C_2) \leq \xi_1' \mathbb{E}_z(C_{i,2}\mu_{\phi,i}\mu'_{\phi,i}C'_{i,2})\xi_1 < \infty$
[9]

[9]For quadratic form $\xi' A\xi$ with A symmetric, $\mathbb{E}\xi' A\xi = \sum_{l=1..\dim(\xi)} \xi_l^2 \mathbb{E}_z a_{ll}^2 + 2\sum_{l,l<k} \xi_l \xi_k \mathbb{E}_z a_{lk} < \sum_l \xi_l^2 \mathbb{E}_z a_{ll}^2 + 2\sum_{l,l<k} \xi_l \xi_k \left(\mathbb{E}_z a_{ll}^2\right)^{\frac{1}{2}} \left(\mathbb{E}_z a_{kk}^2\right)^{\frac{1}{2}}$. Since $\dim\xi < \infty$and for $\hat{\Xi}_{23}$, $a_{ll} = b_{i,ll}^2 \mu_{i,ll}^2$, then $\mathbb{E}_z a_{ll}^2 < (\mathbb{E}_z b_{i,ll}^4)^{\frac{1}{2}} (\mathbb{E}_z \mu_{i,ll}^4)^{\frac{1}{2}} \leq$

Similar argument is followed for $\hat{\Xi}_{21}$.

Combining the above results we can see that:

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

and therefore

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

Below we derive the exact form of the Ω. Under correct specification, by a WLLN, all of the averaged quantities below converge pointwise to some non random function. Furthermore, given that all of these quantities are functions of $m(x,z)$, $M(x,z)$, $D(x,z)$ using measure F or P, we can bound them by dominating functions by taking the supremum over Ψ. By assumption $\mathbf{BD}1-a$ they are bounded. Therefore we can establish uniform convergence to some non random limit.

Recall that we have the system of first order conditions, $G_n(\hat{\vartheta},\hat{\varphi}) = 0$. By the mean value theorem we know the following:

$$0 = \quad G_n(\psi_0) + G_{\psi,n}(\tilde{\psi})(\psi - \psi_0)$$

We have already established a CLT for $G_n(\psi_0)$.

With regard to $G_{\psi,n}(\tilde{\psi}) \equiv \nabla_{(\vartheta,\phi)}$, recall that the population Jacobian matrix is the following:

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

where

$$
\begin{aligned}
G_{i,\vartheta_l\vartheta'} &= m_i(\vartheta)'\mu_{i,\vartheta_l\vartheta'} + \mu'_{i,\vartheta'}(M_i(\vartheta) - \frac{1}{N_s}\sum_j^{N_s}\tilde{e}_jM_j(\vartheta)) + M_i(\vartheta)'\mu_{i,\vartheta_l} + \ldots \\
&\ldots + \mu'_i(D_{i,l,\vartheta'}(\vartheta) - \frac{1}{N_s}\sum_j^{N_s}\tilde{e}_jD_{j,l,\vartheta'}(\vartheta)) - \mu'_i\frac{1}{N_s}\sum_j^{N_s}\tilde{e}_{j,\vartheta'}M_j(\vartheta) \\
\bar{G}_{i,\vartheta_l\vartheta'} &\underset{p}{\rightarrow} \mathbb{E}M_i(\vartheta)'V_m^{-1}(\vartheta)M_{i,l}(\vartheta) \\
G_{i,\vartheta_l\varphi'} &= (M_{i,l}(\vartheta)' - \frac{1}{N_s}\sum_j^{N_s}\tilde{e}_jM_{j,l}(\vartheta)')\mu_{i,\varphi'} + m_i(\vartheta)'\mu_{i,\vartheta_l\varphi'} + \ldots \\
&\ldots - \mu'_i(\frac{1}{N_s}\sum_j^{N_s}\tilde{e}_jM_{j,l}(\vartheta)\mathfrak{s}_j(\varphi) - \frac{1}{N_s}\sum_j^{N_s}\tilde{e}_j\mathfrak{s}_j(\varphi)\frac{1}{N_s}\sum_j^{N_s}\tilde{e}_jM_j(\varphi)) + \ldots \\
&\ldots - (\sum_j^{N_s}\tilde{e}_jm_j(\vartheta)'\mathfrak{s}_j(\varphi))\mu_\vartheta
\end{aligned}
$$

$(\mathbb{E}_z\mathbb{E}_{f|z}M(\vartheta)^4_{i,ll})^{\frac{1}{2}}(\mathbb{E}_z\mathbb{E}_{f|z}m(\vartheta)^4_{i,ll})^{\frac{1}{2}} < \infty$. Similarly, for $\hat{\Xi}_{25}$, $a_{ll} = c^2_{i,ll}\mu^2_{\phi,i,ll}$, then $\mathbb{E}_z a^2_{ll} < (\mathbb{E}_z c^4_{i,ll})^{\frac{1}{2}}(\mathbb{E}_z\mu^4_{\phi,i,ll})^{\frac{1}{2}} \leq (\mathbb{E}_z\mathbb{E}_{f|z}(m(\vartheta))^4_{i,ll})^{\frac{1}{2}}(\mathbb{E}_z\mathbb{E}_{f|z}\nabla_\phi f(\vartheta)^4_{i,ll})^{\frac{1}{2}} < \infty$ Consequently, $\mathbb{E}\xi'A\xi < \infty$

$$\begin{aligned}
\bar{G}_{i,\vartheta_l\varphi'} &\underset{p}{\rightarrow} \mathbb{E}_z M(\vartheta) V_m^{-1} \mathbb{E} m_i(\vartheta)' \mathfrak{s}_j(\varphi) \\
G_{i,\varphi_l\varphi'} &= \frac{\partial^2 log(f_i(\varphi))}{\partial \varphi_l \varphi'} - \frac{1}{N_S}\sum_j \tilde{e}_j \frac{\partial^2 log(f_{j,i}(\varphi))}{\partial \varphi_l \varphi'} + m_i(\vartheta)'\mu_{i,\varphi_l\varphi'} + \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{l,j} \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_j' \\
\bar{G}_{i,\varphi_l\varphi'} &\underset{p}{\rightarrow} \mathbb{E}_z \mathfrak{s}_{l,i} \mathfrak{s}_{l,i}'
\end{aligned}$$

Let $\Lambda_i \equiv M_i(\vartheta) - \sum_j \tilde{e}_j M_{j,i}(\vartheta)$

For $V_{f,m} = \frac{1}{N_s}\sum_j^{N_s} m_j(\vartheta) m_j(\vartheta)'$, let $V_{f,m}^{-1} \equiv (v_1', v_s'..v_{n_m}')'$ and $v_1 \equiv (v_{11}, v_{12}...v_{1n_m})'$. Notice also that for any expansion that follows, we use the fact that for $\bar{V}_{f,m} \equiv \mathbb{E}_{y_s} V_{f,m}$

$$\begin{aligned}
V_{f,m}^{-1} &= \bar{V}_{f,m}^{-1} - \bar{V}_{f,m}^{-1}(V_{f,m} - \bar{V}_{f,m})\bar{V}_{f,m}^{-1} + O_p(N_s^{-1}) \\
&= \bar{V}_{f,m}^{-1} + O_p(N_s^{-\frac{1}{2}})
\end{aligned}$$

or more generally, for some function $g : ||f-g|| < C_n$:

$$\bar{V}_{g,m}^{-1} = \bar{V}_{f,m}^{-1} - \bar{V}_{f,m}^{-1}(\bar{V}_{g,m} - \bar{V}_{f,m})\bar{V}_{f,m}^{-1} + O_p(||g-f||_{TV}^2)$$

Therefore,

$$\begin{aligned}
\bar{V}_{g,m}^{-1} &= \bar{V}_{f,m}^{-1} - \bar{V}_{f,m}^{-1}(\bar{V}_{g,m} - \bar{V}_{f,m})\bar{V}_{f,m}^{-1} + O_p(||g-f||_{TV}^2) \\
&= \bar{V}_{f,m}^{-1} - \bar{V}_{f,m}^{-1}(\bar{V}_{g,m} - \bar{V}_{f,m})\bar{V}_{f,m}^{-1} + O_p(||g-f||_{TV}^2) \\
&= \bar{V}_{f,m}^{-1} + \kappa_n^{-1}\Delta_v(Z)
\end{aligned}$$

With regard to variance covariance matrices computed under the perturbed measure,

$$\begin{aligned}
V_{e,f,m} &= \int (e-1) m(\vartheta) m(\vartheta)' f(x|z) dx + \int m(\vartheta) m(\vartheta)' f(x|z) dx \\
&= \int (m(\vartheta)'(\mu - \mu_0)) m(\vartheta) m(\vartheta)' f(x|z) dx + \int m(\vartheta) m(\vartheta)' f(x|z) dx + O_p(\kappa_n^{-2}) \\
&= \kappa_n^{-1} \int m(\vartheta)'\delta(z) m(\vartheta) m(\vartheta)' f(x|z) dx + \int m(\vartheta) m(\vartheta)' f(x|z) dx + O_p(\kappa_n^{-2}) \\
&= \int m(\vartheta) m(\vartheta)' f(x|z) dx + \kappa_n^{-1}\Delta(z) + O_p(\kappa_n^{-2}) \\
&= V_{f,m} + \kappa_n^{-1}\Delta(z) + O_p(\kappa_n^{-2})
\end{aligned}$$

Also, for any $(y,z)-$ measurable vector or matrix $S(y,z)$, a generic element of the product with $V_{f,m}^{-1}$ is as follows:

$$\begin{aligned}
\mathbb{E}V_{*f,m} &= \mathbb{E}V^{-1}_{*1,f,m}\frac{1}{N_s}\sum_j S(y_j,z_j)\\
&= \sum_{l=1}^{N_m}\mathbb{E}V^{-1}_{*f,m,(1,l)}\frac{1}{N_s}\sum_j S(y_j,z_j)_{(l,1)}\\
&= \sum_{l=1}^{N_m}\bar{V}^{-1}_{*f,m,(1,l)}\mathbb{E}S(y,z)_{(l,1)} - \frac{1}{N_s}\sum_{l=1}^{N_m}\mathbb{C}ov((\bar{V}^{-1}_{f,m,(1,l)}V_{f,m,(1,l)}\bar{V}^{-1}_{f,m,(1,l)} - \bar{V}^{-1}_{f,m,(1,l)})S(y_j,z_j)_{(l,1)}) + \ldots\\
&\quad \ldots + O_p(N_s^{-2})
\end{aligned}$$

In the following derivations we directly use $\bar{V}^{-1}_{f,m}$ for $V^{-1}_{f,m}$ as any remainder will be of even higher order.We derive below the variance covariance matrix:

$$\begin{aligned}
\mathbb{V}(\kappa_n\hat{g}_1(\vartheta)) &= \left(\frac{\kappa_n}{N_0}\right)^2(\underbrace{\mathbb{E}(\sum_i\hat{\mu}_i'\Lambda_i + \sum_i m_i'\hat{\mu}_{i,\vartheta})'(\sum_i\hat{\mu}_i'\Lambda_i + \sum_i m_i'\hat{\mu}_{i,\vartheta})}_{\Gamma}\\
&\quad - \left(\kappa_n^{-2}\sum_l^{n_m}\int\delta(Z)M_{i,(1,l)}(\vartheta)\delta(Z)_{\mu_l}d(Y,Z) - \frac{1}{N_s}\sum_l^{n_m}\mathbb{E}_{y_s}\left(\tilde{\Delta}_{j,l}\Delta_{j,l}\right)\right)^2)
\end{aligned}$$

For $\Gamma \equiv \Gamma_1 + \Gamma_2 + \Gamma_2' + \Gamma_3$:

$$\begin{aligned}
\mathbb{E}\Gamma_1 &= \sum_i\mathbb{E}\Lambda_i'\hat{\mu}_i\hat{\mu}_i'\Lambda_i + 2\sum_{i>j}\mathbb{E}\Lambda_j'\hat{\mu}_j\hat{\mu}_i'\Lambda_i\\
\mathbb{E}\Gamma_2 &= \sum_i\mathbb{E}\hat{\mu}_{i,\vartheta}'m_i\hat{\mu}_i'\Lambda_i + 2\sum_{i>j}\mathbb{E}\hat{\mu}_{j,\vartheta}'m_j\hat{\mu}_i'\Lambda_i\\
\mathbb{E}\Gamma_3 &= \sum_i\mathbb{E}\hat{\mu}_{i,\vartheta}'m_im_i'\hat{\mu}_{i,\vartheta} + 2\sum_{i>j}\mathbb{E}\hat{\mu}_{j,\vartheta}'m_jm_i'\hat{\mu}_{i,\vartheta}
\end{aligned}$$

Using a similar type of expansion for $\hat{\mu}$ as for $\hat{\vartheta} - \vartheta_0$,

$$\mu - \mu_{i,0} = -N_s^{-\frac{1}{2}}\bar{Q}^{-1}_{\mu\mu}\tilde{Q}_\mu + O_p(N_s^{-1})$$

Let also $v_* = M'V^{-1}_{f.m}$. Therefore,

$$\begin{aligned}
\Lambda_i'\hat{\mu}_i\hat{\mu}_i'\Lambda_i &= (M_i(\vartheta) - \frac{1}{N_s}\sum_j M_j(\vartheta))'\frac{1}{N_s}\sum_j\bar{V}'_{*j}m_j(\vartheta)\frac{1}{N_s}\sum_j m_j(\vartheta)'\bar{V}_{*j}(M_i(\vartheta) - \frac{1}{N_s}\sum_j M_j(\vartheta))'\\
&= M_i(\vartheta)'\frac{1}{N_s}\sum_j\bar{V}'_{*j}m_j(\vartheta)\frac{1}{N_s}\sum_j V_{*j}m_j(\vartheta)'M_i(\vartheta)'\ldots
\end{aligned}$$

$$
\begin{aligned}
&\ldots - M_i(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'^{-1}_{*j} m_j(\vartheta)\frac{1}{N_s}\sum_j \bar{V}_{*j} m_j(\vartheta)'\frac{1}{N_s}\sum_j M_j(\vartheta)' + \ldots \\
&\ldots - \frac{1}{N_s}\sum_j M_j(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'_{*j} m_j(\vartheta)\frac{1}{N_s}\sum_j m_j(\vartheta)'\bar{V}_{*j} M_i(\vartheta)' \\
&+\frac{1}{N_s}\sum_j M_j(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'_{*j} m_j(\vartheta)\frac{1}{N_s}\sum_j m_j(\vartheta)'\bar{V}_{*j}\frac{1}{N_s}\sum_j M_j(\vartheta)' \\
= \;& \mathbb{E}\Xi_1 + \mathbb{E}\Xi_2 + \mathbb{E}\Xi_2' + \mathbb{E}\Xi_3
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\Xi_1 &= \frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta) V_{*j} m_j(\vartheta)'|Z) M_i(\vartheta)' + \ldots \\
&\quad \ldots + \frac{N_s-1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j} m_j(\vartheta)'|Z) M_i(\vartheta)' \\
&= \frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta) V_{*j} m_j(\vartheta)'|Z) M_i(\vartheta)' \\
&\quad -\frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j} m_j(\vartheta)'|Z) M_i(\vartheta)' \\
&\quad \ldots + \mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j} m_j(\vartheta)'|Z) M_i(\vartheta)'
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\Xi_2 &= \ldots - O_p(N_s^{-2}) - \frac{N_s-1}{N_s^2}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)\bar{V}_{*j} m_j(\vartheta)'|Z)\mathbb{E}_{y_s}(M_j(\vartheta)'|Z)\ldots \\
&\quad \ldots. - \frac{N_s-1}{N_s^2}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{*j} m_j(\vartheta)' M_j(\vartheta)'|Z)\ldots \\
&\quad \ldots. - \frac{(N_s-1)(N_s-2)}{3!N_s^2}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{*j} m_j(\vartheta)'|Z)\mathbb{E}_{y_s}(M_j(\vartheta)'|Z) \\
&= \ldots - O_p(N_s^{-2}) - \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)\bar{V}_{*j} m_j(\vartheta)'|Z)\mathbb{E}_{y_s}(M_j(\vartheta)'|Z)\ldots \\
&\quad \ldots - \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{*j} m_j(\vartheta)' M_j(\vartheta)'|Z). \\
&\quad -\frac{(N_s-1)(N_s-2)}{3!N_s^2}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'^{-1}_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{*j} m_j(\vartheta)'|Z)\mathbb{E}_{y_s}(M_j(\vartheta)'|Z)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\Xi_3 &= \frac{1}{N_s}\mathbb{E}\sum_j M_j(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'_{*j} m_j(\vartheta)\frac{1}{N_s}\sum_j m_j(\vartheta)'\bar{V}_{*j}\frac{1}{N_s}\sum_j M_j(\vartheta) \\
&= O_p(N_s^{-2}) + \mathbb{E}\frac{(N_s-1)(N_s-2)}{3!N_s^3}(\mathbb{E}_{y_s}(M_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta) m_j(\vartheta)'\bar{V}_{*j} M_j(\vartheta)|Z) + \ldots \\
&\quad \ldots + \mathbb{E}_{y_s}(M_j(\vartheta)\bar{V}'_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(m_j(\vartheta)'\bar{V}_{*j} M_j(\vartheta)|Z) + \ldots \\
&\quad \ldots + \mathbb{E}_{y_s}(M_j(\vartheta)\bar{V}'_{*j} m_j(\vartheta) m_j(\vartheta)'\bar{V}_{*j}|Z)\mathbb{E}_{y_s}(M_j(\vartheta)|Z)) + \ldots \\
&\quad \ldots + \frac{(N-1)(N-2)(N-3)}{4!N^3}\mathbb{E}_{y_s}(M_j(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}'_{*j} m_j(\vartheta)|Z)\mathbb{E}_{y_s}(m_j(\vartheta)'\bar{V}_{*j}|Z)\mathbb{E}_{y_s}(M_j(\vartheta)|Z)
\end{aligned}
$$

With regard to the lagged terms, we show below an indicative derivation, which is similare to the variance derivation, so we skip the rest for the sake of brevity.

$$\begin{aligned}
\Lambda_i'\hat{\mu}_i\hat{\mu}_l'\Lambda_l &= (M_i(\vartheta)-\frac{1}{N_s}\sum_j M_{j,i}(\vartheta))'\frac{1}{N_s}\sum_j \bar{V}_{*j,i}'m_{j,i}(\vartheta)\frac{1}{N_s}\sum_j m_{j,l}(\vartheta)'\bar{V}_{*j,l}(M_l(\vartheta)-\frac{1}{N_s}\sum_j M_{j,l}(\vartheta))' \\
&= M_i(\vartheta)'\frac{1}{N_s}\sum_j \bar{V}'_{*j,i}m_{j,i}(\vartheta)\frac{1}{N_s}\sum_j V_{*j,l}m_{j,l}(\vartheta)'M_l(\vartheta)'... \\
&\quad ... - M_i(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'^{-1}_{*j,i}m_{j,i}(\vartheta)\frac{1}{N_s}\sum_j \bar{V}_{*j,l}m_{j,l}(\vartheta)'\frac{1}{N_s}\sum_j M_{j,l}(\vartheta)'+... \\
&\quad ... - \frac{1}{N_s}\sum_j M_{j,i}(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'_{*j,i}m_{j,i}(\vartheta)\frac{1}{N_s}\sum_j m_{j,l}(\vartheta)'\bar{V}_{*j,l}M_l(\vartheta)' \\
&\quad + \frac{1}{N_s}\sum_j M_{j,i}(\vartheta)\frac{1}{N_s}\sum_j \bar{V}'_{*j,i}m_{j,i}(\vartheta)\frac{1}{N_s}\sum_j m_{j,l}(\vartheta)'\bar{V}_{*j,l}\frac{1}{N_s}\sum_j M_{j,l}(\vartheta)' \\
&= \mathbb{E}\Xi_1+\mathbb{E}\Xi_2+\mathbb{E}\tilde{\Xi}_2+\mathbb{E}\Xi_3
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Xi_1 &= \frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j,i}m_{j,i}(\vartheta)V_{*j,l}m_{j,l}(\vartheta)'|Z)M_l(\vartheta)'+... \\
&\quad ... + \frac{N_s-1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j,i}m_{j,i}(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j,l}m_{j,l}(\vartheta)'|Z)M_{i,l}(\vartheta)' \\
&= \frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j,i}m_{j,i}(\vartheta)V_{*j,l}m_{j,l}(\vartheta)'|Z)M_l(\vartheta)' \\
&\quad - \frac{1}{N_s}\mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j,i}m_j(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j,l}m_{j,l}(\vartheta)'|Z)M_l(\vartheta)' \\
&\quad ... + \mathbb{E}_z M_i(\vartheta)'\mathbb{E}_{y_s}(\bar{V}'_{*j}m_{j,i}(\vartheta)|Z)\mathbb{E}_{y_s}(V_{*j,l}m_{j,l}(\vartheta)'|Z)M_l(\vartheta)'
\end{aligned}$$

With regard to the second term,

$$\begin{aligned}
\mathbb{E}\hat{\mu}_{i,\vartheta}'m_i\hat{\mu}_i'\Lambda_i &= -\mathbb{E}\frac{1}{N_s}\sum_j(M_{j,i}(\vartheta)'V_{e*j,i}z_j')m_i(\vartheta)\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)'V_{*j,i}'(M_i(\vartheta)-\frac{1}{N_s}\sum_j M_{j,i}(\vartheta)) \\
&= -\mathbb{E}\frac{1}{N_s}\sum_j(M_{j,i}(\vartheta)'V_{e*j,i}z_j')m_i(\vartheta)\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)V_{*j,i}'M_i(\vartheta) \\
&\quad +\mathbb{E}\frac{1}{N_s}\sum_j(M_{j,i}(\vartheta)'V_{e*j,i}z_j')m_i(\vartheta)\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)V_{*j,i}'\frac{1}{N_s}\sum_j M_{j,i}(\vartheta) \\
&= -\frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'(V_{e*j,i}z_j')\mathbb{E}_y(m_i(\vartheta)m_{j,i}(\vartheta)V_{*j,i}'|Z)M_i(\vartheta)|Z) \\
&\quad -\frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'V_{e*j,i}z_j'|Z)\mathbb{E}_y(m_i(\vartheta)\mathbb{E}_{y_s}(m_{j,i}(\vartheta)V_{*j,i}'|Z)M_i(\vartheta)|Z) \\
&\quad +O_p(N_s^{-2})+\frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'V_{e*j,i}z_j'|Z)\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}_{y_s}(m_{j,i}(\vartheta)V_{*j,i}'M_{j,i}(\vartheta)|Z) \\
&\quad +\frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'V_{e*j,i}z_j'\mathbb{E}_y(m_i(\vartheta)|Z)m_{j,i}(\vartheta)V_{*j,i}'|Z)\mathbb{E}_{y_s}(M_{j,i}(\vartheta)|Z)
\end{aligned}$$

$$+\frac{(N-1)(N-2)}{3!N^2}\mathbb{E}_z\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'V_{e*j,i}z_j'|Z)\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}_{y_s}(m_{j,i}(\vartheta)V_{*j,i}'|Z)\mathbb{E}_{y_s}(M_{j,i}(\vartheta)|Z)$$

With regard to the third term,

$$\begin{aligned}
\mathbb{E}\hat{\mu}_{i,\vartheta}'m_im_i'\hat{\mu}_{i,\vartheta} &= \mathbb{E}\frac{1}{N_s}\sum_j M_{j,i}(\vartheta)'V_{e,j,i}^{-1}z_jm_im_i'\frac{1}{N_s}\sum_j V_{e,j,i}^{-1}z_jM_{j,i}(\vartheta)\\
&= \mathbb{E}_z\frac{1}{N_s}\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}^{-1}z_j\mathbb{E}_y(m_im_i'|Z)V_{e,j,i}^{-1}z_jM_{j,i}(\vartheta)+\ldots.\\
&\quad+\mathbb{E}_z\frac{N_s-1}{N_s}\mathbb{E}_{y_s}(M_{j,i}(\vartheta)'V_{e,j,i}^{-1}z_j)\mathbb{E}_y(m_im_i'|Z)\mathbb{E}_{y_s}V_{e,k,i}^{-1}z_kM_{k,i}(\vartheta)|Z)\\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}z_j\mathbb{E}_y(m_im_i'|Z)V_{e,j,i}z_jM_{j,i}(\vartheta)+\ldots.\\
&\quad+\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)-\frac{1}{N_s}\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)+\ldots\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_{j,i}(\vartheta)'\mathbb{E}V_{e,j,i}^{-1}z_j-\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)^{-1})\mathbb{E}(V_i|Z)(\mathbb{E}_{y_s}V_{e,k,i}z_jM_{k,i}(\vartheta)\\
&\quad+\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}(\mathbb{E}(M_i(\vartheta)|Z)-M_{k,i}(\vartheta))'V_{e,k,i}z_kM_{k,i}(\vartheta))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}M_{k,i}(\vartheta)'V_{e,k,i}z_k(M_{k,i}(\vartheta)-\mathbb{E}(M_i(\vartheta)|Z))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_{z,y_s}(M_{k,i}(\vartheta)'V_{e,k,i}z_k-M_{j,i}(\vartheta)'V_{e,j,i}z_j)\\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}z_j\mathbb{E}_y(m_im_i'|Z)V_{e,j,i}z_jM_{j,i}(\vartheta)+\ldots.\\
&\quad+\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)-\frac{1}{N_s}\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)+\ldots\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_{j,i}(\vartheta)'(\kappa_n^{-1}\Delta(Z)+\mathbb{E}_{y_s}V_{f,m}+O_p(\kappa_n^{-2}))\\
&\quad-\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)^{-1})\mathbb{E}(V_i|Z)(\mathbb{E}_{y_s}V_{e,k,i}z_jM_{k,i}(\vartheta)+\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(\mathbb{E}(M_i(\vartheta)|Z)-M_{k,i}(\vartheta))'V_{e,k,i}z_kM_{k,i}(\vartheta))+\\
&\quad\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}M_{k,i}(\vartheta)'V_{e,k,i}z_k(M_{k,i}(\vartheta)-\mathbb{E}(M_i(\vartheta)|Z))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_{z,y_s}(M_{k,i}(\vartheta)'V_{e,k,i}z_k-M_{j,i}(\vartheta)'V_{e,j,i}z_j)\\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}z_j\mathbb{E}_y(m_im_i'|Z)V_{e,j,i}z_jM_{j,i}(\vartheta)+\ldots.\\
&\quad+\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)-\frac{1}{N_s}\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z)+\ldots\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_{j,i}(\vartheta)'-\mathbb{E}(M_i(\vartheta)'|Z)(\mathbb{E}(V_i|Z)^{-1})+\\
&\quad\kappa_n^{-1}\Delta_v(Z))\mathbb{E}(V_i|Z)^{-1})\mathbb{E}(V_i|Z)(\mathbb{E}_{y_s}V_{e,k,i}z_jM_{k,i}(\vartheta)+\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z))\\
&= \frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}(\mathbb{E}(M_i(\vartheta)|Z)-M_{k,i}(\vartheta))'(\mathbb{E}(V_i|Z)^{-1})+\kappa_n^{-1}\Delta_v(Z))M_{k,i}(\vartheta))\\
&\quad+\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}M_{k,i}(\vartheta)'V_{e,k,i}z_k(M_{k,i}(\vartheta)-\mathbb{E}(M_i(\vartheta)|Z))
\end{aligned}$$

$$
\begin{aligned}
&+\frac{N_s-1}{N_s}\mathbb{E}_{z,y_s}(M_{k,i}(\vartheta)'V_{e,k,i}z_k - M_{j,i}(\vartheta)'(\mathbb{E}(V_i|Z)^{-1}) + \kappa_n^{-1}\Delta_v(Z)))\\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}z_j\mathbb{E}_y(m_i m_i'|Z)V_{e,j,i}z_jM_{j,i}(\vartheta) + ...\\
&+\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z) - \frac{1}{N_s}\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z) + ...\\
&+\frac{N_s-1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_{j,i}(\vartheta)' - \mathbb{E}(M_i(\vartheta)'|Z)(\mathbb{E}(V_i|Z)^{-1}) +\\
&\kappa_n^{-1}\Delta_v(Z))\mathbb{E}(V_i|Z)^{-1})\mathbb{E}(V_i|Z)(\mathbb{E}_{y_s}V_{e,k,i}z_jM_{k,i}(\vartheta) + \mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z))\\
&+\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}(\mathbb{E}(M_i(\vartheta)|Z) - M_{k,i}(\vartheta))'(\mathbb{E}(V_i|Z)^{-1}) + \kappa_n^{-1}\Delta_v(Z))M_{k,i}(\vartheta)) +\\
&\frac{N_s-1}{N_s}\mathbb{E}_{(y,z)}\mathbb{E}_{y_s}M_{k,i}(\vartheta)'V_{e,k,i}z_k(M_{k,i}(\vartheta) - \mathbb{E}(M_i(\vartheta)|Z))\\
&+\frac{N_s-1}{N_s}\mathbb{E}_{z,y_s}(M_{k,i}(\vartheta)'V_{e,k,i}z_k - M_{j,i}(\vartheta)'(\mathbb{E}(V_i|Z)^{-1}) + \kappa_n^{-1}\Delta_v(Z)))\\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}M_{j,i}(\vartheta)'V_{e,j,i}z_j\mathbb{E}_y(m_i m_i'|Z)V_{e,j,i}z_jM_{j,i}(\vartheta) + ...\\
&+\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z) - \frac{1}{N_s}\mathbb{E}_z\mathbb{E}(M_i(\vartheta)'|Z)\mathbb{E}(V_i|Z)\mathbb{E}(M_i(\vartheta)|Z) + O_p(\kappa_n^{-1})
\end{aligned}
$$

Let

$$
\bar{g}_2(\vartheta) \equiv \frac{1}{N}\sum_i(\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j\tilde{e}_j\mathfrak{s}_{j,i}(\phi)) - \frac{1}{N_s}\sum_j(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))\frac{1}{N}\sum_i V_{m,f}^{-1}m_i(\vartheta)
$$

and denote by $P_{s,y}^{m}$ the projection matrix for projecting the score $\mathfrak{s}(y,z)$ on the moment conditions $m(y,z)$ and $\mathfrak{s}^p(y,z) \equiv m(y,z)P_{s,y}^{m}$ the predicted score. The variance component for the second set of moment conditions is therefore:

$$
\mathbb{V}(\kappa_n\hat{g}_2(\psi)) = \left(\frac{\kappa_n}{N_0}\right)^2(\sum_i\mathbb{E}g_{2,i}(\psi)g_{2,i}(\psi)' + 2\sum_{i>j}\mathbb{E}g_{2,i}(\psi)g_{2,k}(\psi)')
$$

Given ergodicity assumptions, the second component is summable. With regard to the first component :

$$
\begin{aligned}
&\mathbb{E}_{(y,z,y_s)}g_{2,i}(\psi)g_{2,i}(\psi)'\\
=\ &\mathbb{E}_{(y,z,y_s)}(\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j\tilde{e}_j\mathfrak{s}_{j,i}(\phi))(\mathfrak{s}_i(\phi)' - \frac{1}{N_s}\sum_j\tilde{e}_j\mathfrak{s}_{j,i}(\phi)') + ...\\
&... - \mathbb{E}_{(y,z,y_s)}(\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j\tilde{e}_j\mathfrak{s}_{j,i}(\phi))(\frac{1}{N_s}\sum_j(m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}m_i(\vartheta)' + ...\\
&... - \mathbb{E}_{(y,z,y_s)}(\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi)V_{m,f}^{-1}m_i(\vartheta)(\mathfrak{s}_i(\phi)' - \frac{1}{N_s}\sum_j\tilde{e}_j\mathfrak{s}_{j,i}(\phi)') + ...\\
&... + \mathbb{E}_{(y,z,y_s)}(\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}m_i(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi)')\\
=\ &\mathbb{E}_{(y,z,y_s)}(\Xi_1 + \Xi_2' + \Xi_2 + \Xi_3)
\end{aligned}
$$

$$\begin{aligned}
\mathbb{E}_{(y,z,y_s)}\Xi_1 &= \mathbb{E}_{(y,z,y_s)}(\mathfrak{s}_i(\phi)\mathfrak{s}_i(\phi)' - \mathfrak{s}_i(\phi)\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi)' + \ldots \\
&\quad \ldots - \frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi)\mathfrak{s}_i(\phi)' + \frac{1}{N_s^2}\sum_{j,k}\tilde{e}_j\mathfrak{s}_{j,i}(\phi)\tilde{e}_k\mathfrak{s}_{k,i}(\phi)' \\
&= \mathbb{E}_{(y,z,y_s)}\mathfrak{s}_i(\phi)\mathfrak{s}_i(\phi)' - \mathbb{E}_z\mathbb{E}_y(\mathfrak{s}_i(\phi)|Z)\mathbb{E}_y(\tilde{e}_j\mathfrak{s}_{j,i}(\phi)'|Z) - \mathbb{E}_z\mathbb{E}_{y_s}(\tilde{e}_j\mathfrak{s}_{j,i}(\phi)|Z)\mathbb{E}_y(\mathfrak{s}_i(\phi)'|Z) + \ldots \\
&\quad \ldots + \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(\tilde{e}_j\mathfrak{s}_{j,i}(\phi)|Z)\mathbb{E}_{y_s}(\tilde{e}_k\mathfrak{s}_{k,i}(\phi)'|Z) + \frac{1}{N_s^2}\sum_{j\neq k}\mathbb{E}_{y_s}\tilde{e}_j\mathfrak{s}_{j,i}(\phi)\mathbb{E}_{y_s}\tilde{e}_k\mathfrak{s}_{k,i}(\phi)' \\
&= \mathbb{E}_{(y,z,y_s)}\mathfrak{s}_i(\phi)\mathfrak{s}_i(\phi)' - \delta_{\mathfrak{s}}\kappa_n^{-1} + \delta_{\mathfrak{s}_s}\frac{N_s-1}{N_s}\kappa_n^{-1}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{(y,z,y_s)}\Xi_2 &= -\mathbb{E}_{(y,z,y_s)}\frac{1}{N_s}\sum_j (m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}m_i(\vartheta))(\mathfrak{s}_i(\phi)' - \frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi)') \\
&= -\frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}|Z)\mathbb{E}_y(m_i(\vartheta)\mathfrak{s}_i(\phi)'|Z) \\
&\quad + \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}\mathbb{E}_z(m_i(\vartheta)|Z)\tilde{e}_j\mathfrak{s}_{j,i}(\phi)'|Z) \\
&\quad + \frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}|Z)\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}_{y_s}(\tilde{e}_j\mathfrak{s}_{j,i}(\phi)'|Z) \\
&= O_p(N_s^{-1}) + \delta_{\mu_\phi}\kappa_n^{-1} + \mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}|Z)\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}_{y_s}(\tilde{e}_j\mathfrak{s}_{j,i}(\phi)'|Z) \\
&= O_p(N_s^{-1}) + \tilde{\delta}_{\mu_\phi}\kappa_n^{-1} + \mathbb{E}_{y,z}(\mathfrak{s}_i^p(\phi)\mathfrak{s}_i(\phi)'|Z)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{(y,z,y_s)}\Xi_3 &= \mathbb{E}_{(y,z,y_s)}(\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}m_i(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}\frac{1}{N_s}\sum_j m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi)') \\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}V_mV_{m,f}^{-1}(m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi)') + \ldots \\
&\quad \ldots + \mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}|Z)V_m\mathbb{E}_{y_s}(V_{m,f}^{-1}(m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi)'|Z) + \ldots \\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_y(m_i(\vartheta)\otimes\mathfrak{s}_i(\phi))V_m^{-1}(m_i(\vartheta)'\otimes\mathfrak{s}_i(\phi)') + \ldots \\
&\quad \ldots + \mathbb{E}_z\mathbb{E}_{y_s}(m_{j,i}(\vartheta)\otimes\mathfrak{s}_{j,i}(\phi))V_{m,f}^{-1}|Z)m_i(\vartheta)m_i(\vartheta)'\mathbb{E}_{y_s}(V_{m,f}^{-1}(m_{j,i}(\vartheta)'\otimes\mathfrak{s}_{j,i}(\phi)'|Z) + \ldots \\
&\quad \ldots + O_p(\kappa_n^{-1}) \\
&= O_p(\max(N_s^{-1},\kappa_n^{-1})) + \mathbb{E}_{(y,z,y_s)}\mathfrak{s}_i^p(\phi)\mathfrak{s}_i^p(\phi)'
\end{aligned}$$

With regard to the covariance term,

$$\mathbb{C}ov(\kappa_n\hat{g}_1(\psi),\kappa_n\hat{g}_2(\psi)) = \left(\frac{\kappa_n}{N_0}\right)^2(\sum_i \mathbb{E}g_{1,i}(\psi)g_{2,i}(\psi)' + 2\sum_{i>j}\mathbb{E}g_{1,i}(\psi)g_{2,k}(\psi)')$$

$$= \left(\frac{\kappa_n}{N_0}\right)^2 \sum_i \mathbb{E} g_{1,i}(\psi) g_{2,i}(\psi)' + 2\left(\frac{\kappa_n}{N_0}\right)^2 \sum_{i>j} \mathbb{E} g_{1,i}(\psi) g_{2,k}(\psi)')$$

With regard to the first term, ...

$$\begin{aligned}
\mathbb{E} g_{1,i}(\psi) g_{2,i}(\psi)' &= \mathbb{E}(\hat{\mu}_i' \Lambda_i + m_i' \hat{\mu}_{i,\vartheta})'(\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi) - m_i(\vartheta)' V_{m,f}^{-1} \frac{1}{N_s}\sum_j (m_{j,i}(\vartheta)' \otimes \mathfrak{s}_{j,i}(\phi)')) \\
&= \mathbb{E}(\Lambda_i \hat{\mu}_i (\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi)) + ... \\
&\quad ... + \mathbb{E}\hat{\mu}_{i,\vartheta}' m_i(\vartheta)(\mathfrak{s}_i(\phi) - \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi))' + ... \\
&\quad ... + \mathbb{E}(\Lambda_i' \hat{\mu}_i (m_i(\vartheta)' V_{m,f}^{-1}(\vartheta) \frac{1}{N_s}\sum_j (\mathfrak{s}_{j,i}(\phi)' \otimes m_{j,i}(\vartheta)')) + ... \\
&\quad ... + \mathbb{E}(\hat{\mu}_{i,\vartheta}' m_i(\vartheta) m_i(\vartheta)' V_{m,f}^{-1}(\vartheta) \frac{1}{N_s}\sum_j \mathfrak{s}_{j,i}(\phi) \otimes m_{j,i}(\vartheta)') \\
&= \mathbb{E}(\Xi_1 + \Xi_3 + \Xi_2 + \Xi_4)
\end{aligned}$$

Treating each term separately, we have the following:

$$\begin{aligned}
\mathbb{E}\Xi_1 &= \mathbb{E}\Lambda_i' \hat{\mu}_i \mathfrak{s}_i(\phi)' - \Lambda_i' \hat{\mu}_i \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi)' \\
&= \mathbb{E}_z \mathbb{E}_{y_s}((M_i(\vartheta) - \frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta))' \hat{\mu}_i | Z)(\mathfrak{s}_i(\phi)'|Z) - \\
&\quad (M_i(\vartheta) - \frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta))' \hat{\mu}_i \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi)' \\
&= \mathbb{E}_z(\mathbb{E}_y(M_i(\vartheta)' \mathfrak{s}_i(\phi)|Z) \mathbb{E}_{y_s}(\hat{\mu}_i|Z) - \mathbb{E}_z \mathbb{E}_{y_s}(\frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta)' \hat{\mu}_i) \mathbb{E}_y(\mathfrak{s}_i(\phi)|Z) + ... \\
&\quad -\mathbb{E}_z \mathbb{E}_y(M_i(\vartheta)|Z)' \mathbb{E}_{y_s} \hat{\mu}_i \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi) - \mathbb{E}_z \mathbb{E}_{y_s} \frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta)' \hat{\mu}_i \frac{1}{N_s}\sum_j \tilde{e}_j \mathfrak{s}_{j,i}(\phi) \\
&= \mathbb{E}\Xi_{11} + \mathbb{E}\Xi_{12} + \mathbb{E}\Xi_{13} + \mathbb{E}\Xi_{14}
\end{aligned}$$

Since $\mu_i = V_{\kappa,m,f}^{-1} \frac{1}{N_s} \sum_j m_j(\vartheta)$ and using results from Lemma (1),

$$\begin{aligned}
\mathbb{E}\Xi_{11} &= \mathbb{E}_z \mathbb{E}_y(M_i(\vartheta)' \mathfrak{s}_i(\phi)|Z) \mathbb{E}_{y_s} V_{\kappa,m,f}^{-1}(m_j(\vartheta)|Z) \\
&= \mathbb{E}_z \mathbb{E}_y(M_i(\vartheta)' \mathfrak{s}_i(\phi)'|Z)(\bar{V}_{f,m}^{-1} + \kappa_n^{-1}\Delta(z) + O_p(N_s^{-\frac{1}{2}})) \mathbb{E}_{y_s}(m_j(\vartheta)|Z) \\
&= \mathbb{E}_z \mathbb{E}_y(M_i(\vartheta)' \mathfrak{s}_i(\phi)'|Z) \kappa_n^{-1} \bar{V}_{f,m}^{-1} \delta_m(Z) + O_p(\kappa_n^{-1} N^{-\frac{2}{d}})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Xi_{12} &= -\mathbb{E}_z\mathbb{E}_{y_s}(\frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta)'\hat{\mu}_i)\mathbb{E}_y(\mathfrak{s}_i(\phi)|Z) \\
&= \frac{1}{N_s}\mathbb{E}_{y_s}\tilde{e}_j M_{j,i}(\vartheta)(\bar{V}_{f,m}^{-1}+\kappa_n^{-1}\Delta(z)+O_p(N_s^{-\frac{1}{2}}))(m_j(\vartheta)|Z))\mathbb{E}_y(\mathfrak{s}_i(\phi)'|Z)+... \\
&\quad ...+\frac{N_s-1}{N_s}\mathbb{E}_{y_s}(\tilde{e}_k M_{k,i}(\vartheta)|Z)(\bar{V}_{f,m}^{-1}+\kappa_n^{-1}\Delta(z)+O_p(N_s^{-\frac{1}{2}}))\mathbb{E}_{y_s}(m_j(\vartheta)|Z)\mathbb{E}_y(\mathfrak{s}_i(\phi)'|Z) \\
&= O_p(\kappa_n^{-2})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Xi_{13} &= -\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)'\mathbb{E}_{y_s}\hat{\mu}_i\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi) \\
&= \frac{1}{N_s}\mathbb{E}_z\mathbb{E}_{y_s}\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{f,m}^{-1}m_j(\vartheta)\tilde{e}_j\mathfrak{s}_{j,i}(\phi)|Z)+... \\
&\quad ..+\frac{N_s-1}{N_s}\mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)|Z)\mathbb{E}_{y_s}(\bar{V}_{f,m}^{-1}m_j(\vartheta)|Z)\mathbb{E}_{y_s}\tilde{e}_j\mathfrak{s}_{j,i}(\phi) \\
&= O_p(\max(N_s^{-1},\kappa_n^{-2}))
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Xi_{14} &= -\mathbb{E}_z\mathbb{E}_{y_s}\frac{1}{N_s}\sum_j \tilde{e}_j M_{j,i}(\vartheta)'\hat{\mu}_i\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi) \\
&= O_p(N_s^{-2})+\frac{(N_s-1)(N_s-2)}{3!N_s^2}\mathbb{E}_{y_s}(M_i(\vartheta)|Z)\mathbb{E}_{y_s}\bar{V}_{f,m}^{-1}m_j(\vartheta)\mathbb{E}_{y_s}\tilde{e}_j\mathfrak{s}_{j,i}(\phi) \\
&= O_p(\max(N_s^{-2},\kappa_n^{-2}))
\end{aligned}$$

With regard to Ξ_2:

$$\begin{aligned}
\mathbb{E}\Xi_2 &= \mathbb{E}\hat{\mu}_{i,\vartheta}'m_i(\vartheta)(\mathfrak{s}_i(\phi)-\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi))' \\
&= \mathbb{E}\hat{\mu}_{i,\vartheta}'m_i(\vartheta)\mathfrak{s}_i(\phi)'-\mathbb{E}\hat{\mu}_{i,\vartheta}'m_i\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi)' \\
&= \mathbb{E}\hat{\mu}_{i,\vartheta}'m_i(\vartheta)\mathfrak{s}_i(\phi)'-\mathbb{E}_z\mathbb{E}_{y_s}(\hat{\mu}_{i,\vartheta}\mathbb{E}_y(m_i'|Z)\frac{1}{N_s}\sum_j \tilde{e}_j\mathfrak{s}_{j,i}(\phi)'|Z) \\
&= \mathbb{E}(\bar{V}_{f,m}^{-1}+O_p(N_s^{-\frac{1}{2}}))m_i(\vartheta)\mathfrak{s}_i(\phi)'+O_p(\kappa_n^{-1}) \\
&= \mathbb{E}_z\mathbb{E}_{y_s}(z_jM_j(\vartheta)'|Z)\bar{V}_{f,m}^{-1}\mathbb{E}_y(m_i(\vartheta)\mathfrak{s}_i(\phi)')+O_p(\max(N_s^{-\frac{1}{2}},\kappa_n^{-1})) \\
&= \mathbb{E}_z\mathbb{E}_y(M_j(\vartheta)'|Z)\bar{V}_m^{-1}\mathbb{E}_y(m_i(\vartheta)\mathfrak{s}_i(\phi)')+O_p(\max(N_s^{-\frac{1}{2}},\kappa_n^{-1}))
\end{aligned}$$

With regard to Ξ_3 :

$$\begin{aligned}
\mathbb{E}\Xi_3 &= \mathbb{E}(\Lambda_i'\hat{\mu}_i m_i(\vartheta)V_{m,f}^{-1}(\vartheta)\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')) \\
&= \mathbb{E}((M_i(\vartheta)-\frac{1}{N_s}\sum_j\tilde{e}_jM_{j,i}(\vartheta))'V_{\kappa m,f}^{-1}\frac{1}{N_s}\sum_j m_j(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}(\vartheta)(\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')) \\
&= \mathbb{E}(M_i(\vartheta)'V_{\kappa m,f}^{-1}\frac{1}{N_s}\sum_j m_j(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}(\vartheta)\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)'))+... \\
&\quad ...-\mathbb{E}_z\mathbb{E}_{y_s}(\frac{1}{N_s}\sum_j\tilde{e}_jM_{j,i}(\vartheta)'V_{\kappa m,f}^{-1}\frac{1}{N_s}\sum_j m_j(\vartheta)\mathbb{E}_y(m_i(\vartheta)|Z)'V_{m,f}^{-1}(\vartheta)(\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')) \\
&= \frac{1}{N_s}\mathbb{E}(M_i(\vartheta)'\mathbb{E}_{y_s}V_{\kappa m,f}^{-1}m_j(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')+... \\
&\quad ...+\frac{N_s-1}{N_s}\mathbb{E}(M_i(\vartheta)'\mathbb{E}_{y_s}(V_{\kappa m,f}^{-1}m_j(\vartheta)|Z)m_i(\vartheta)'\mathbb{E}_{y_s}(V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')|Z)+ \\
&\quad ...-\mathbb{E}_z\frac{N_s-1}{N_s^2}\mathbb{E}_{y_s}\tilde{e}_jM_{j,i}(\vartheta)'V_{\kappa m,f}^{-1}m_j(\vartheta)|Z)\mathbb{E}_y(m_i(\vartheta)|Z)'\mathbb{E}_{y_s}V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')|Z)+O_p(N_s^{-2}) \\
&\quad ....-\mathbb{E}_z\frac{N_s-1}{N_s^2}\mathbb{E}_{y_s}(\tilde{e}_jM_{j,i}(\vartheta)'|Z)V_{\kappa m,f}^{-1}\mathbb{E}(m_j(\vartheta)|Z)\mathbb{E}_y(m_i(\vartheta)|Z)'V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')|Z) \\
&\quad ....-\mathbb{E}_z\frac{N_s-1}{N_s^2}\mathbb{E}_{y_s}(\tilde{e}_jM_{j,i}(\vartheta)'V_{\kappa m,f}^{-1}m_j(\vartheta)\mathbb{E}_y(m_i(\vartheta)|Z)'V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)'\otimes m_{j,i}(\vartheta)')|Z) \\
&= \mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)'|Z)V_m^{-1}\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}(m_i(\vartheta)'|Z)V_m^{-1}\mathbb{E}_y(\mathfrak{s}_i(\phi)'\otimes m_i(\vartheta)')|Z)+O_p(\max(N_s^{-\frac{1}{2}},\kappa_n^{-1}) \\
&= \mathbb{E}_z\mathbb{E}_y(M_i(\vartheta)'|Z)V_m^{-1}\mathbb{E}_y(m_i(\vartheta)|Z)\mathbb{E}_y(\mathfrak{s}^{\mathfrak{p}}{}_i(\phi)|Z)+O_p(\max(N_s^{-\frac{1}{2}},\kappa_n^{-1})
\end{aligned}$$

Finally, with regard to Ξ_4:

$$\begin{aligned}
\mathbb{E}\Xi_4 &= \mathbb{E}(\hat{\mu}_{i,\vartheta}'m_i(\vartheta)m_i(\vartheta)'V_{m,f}^{-1}(\vartheta)\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)\otimes m_{j,i}(\vartheta)') \\
&= \mathbb{E}(\frac{1}{N_s}\sum_j z_jM_j(\vartheta)'(\bar{V}_{f,m}^{-1}+O_p(N_s^{-\frac{1}{2}})m_i(\vartheta)m_i(\vartheta)'(V_{m,f}^{-1}(\vartheta)+O_p(N_s^{-\frac{1}{2}}))(\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)\otimes m_{j,i}(\vartheta))) \\
&= \mathbb{E}(\frac{1}{N_s}\sum_j z_jM_j(\vartheta)'\bar{V}_{f,m}^{-1}m_i(\vartheta)m_i(\vartheta)'\bar{V}_{m,f}^{-1}(\vartheta)(\frac{1}{N_s}\sum_j(\mathfrak{s}_{j,i}(\phi)\otimes m_{j,i}(\vartheta))))+... \\
&\quad ...+o_p(N_s^{-\frac{1}{2}}) \\
&= \frac{1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_j(\vartheta)'\bar{V}_{f,m}^{-1}\mathbb{E}_y(m_i(\vartheta)m_i(\vartheta)'|Z)V_{m,f}^{-1}(\vartheta)(\mathfrak{s}_{j,i}(\phi)\otimes m_{j,i}(\vartheta))+... \\
&\quad ...+\frac{N_s-1}{N_s}\mathbb{E}_z(\mathbb{E}_{y_s}M_j(\vartheta)'\bar{V}_{f,m}^{-1}\mathbb{E}_y(m_i(\vartheta)m_i(\vartheta)'|Z)V_{m,f}^{-1}(\vartheta)\mathbb{E}_{y_s}(\mathfrak{s}_{k,i}(\phi)\otimes m_{k,i}(\vartheta)) \\
&\quad ...+o_p(N_s^{-\frac{1}{2}}) \\
&= \mathbb{E}_z(\mathbb{E}_{y_s}M_j(\vartheta)'\bar{V}_{f,m}^{-1}\mathbb{E}_y(m_i(\vartheta)m_i(\vartheta)'|Z)V_{m,f}^{-1}\mathbb{E}_{y_s}(\mathfrak{s}_{k,i}(\phi)\otimes m_{k,i}(\vartheta))+O_p(N_s^{-1})+o_p(N_s^{-\frac{1}{2}}) \\
&= \mathbb{E}_z(\mathbb{E}_yM_{j,i}(\vartheta)'\bar{V}_m^{-1}\mathbb{E}_y(m_i(\vartheta)m_i(\vartheta)'|Z)V_m^{-1}\mathbb{E}_y(\mathfrak{s}_i(\phi)\otimes m_i(\vartheta))+O_p(max(\kappa_n^{-1},N_s^{-1}))+o_p(N_s^{-\frac{1}{2}}) \\
&= \mathbb{E}_z(\mathbb{E}_yM_{j,i}(\vartheta)'V_m^{-1}\mathbb{E}_y(\mathfrak{s}_i(\phi)\otimes m_i(\vartheta))+O_p(max(\kappa_n^{-1},N_s^{-1}))+o_p(N_s^{-\frac{1}{2}})
\end{aligned}$$

□

Proof. of Proposition 7

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

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

□

Proof. of Proposition 8

Substituting the result of Proposition 1 in 3.12 we get that:

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

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:

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

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

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

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

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:

$$tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r)) < n_1 \tag{3.23}$$

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 3.23 holds for any $R \neq 0$. Towards this, we first rewrite the left hand side of 3.23 as follows:

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

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

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

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

What needs to be shown is the following:

$$\max_V \quad tr((V'(J)BV(J))^{-1}V(J)'AV(J)) \quad = n_1 \tag{3.24}$$

The problem defined by the LHS is of 3.24 is a well defined problem in discriminant analysis for a *general matrix* V, and is equivalent to:

$$\max_V \quad tr(V(J)'AV(J))$$
$$\text{s.t} \quad V'(J)BV(J) < K$$

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

$$AV(J) = BV(J)\Lambda \tag{3.25}$$

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

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

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

Since the system of equations in 3.25 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 3.25 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 $R \neq 0$, $tr(\mathbb{V}(S_1\mathcal{Z})^{-1}\mathbb{V}(\mathcal{Z}_r)) < n_1$. □

The following lemmata are systematically used in the proofs of the propositions above:

Lemma 4. *For any $\mathcal{Z}$-measurable function $g(\mu)$, we have that $\mathbb{E}_z g(\hat{\mu}_i) \to \mathbb{E}_z g(\mu_i)$. Consequently, $\mathbb{E}_z\hat{\lambda}_i \to \mathbb{E}_z\lambda$*

Proof. Given results above, consider that (a) $||\hat{\mu}_i - \mu_i|| < C^{-1}|Q'_\mu(\mu_0, z_i)|$. By assumption (**BD** $-$ **1b**) the **RHS** is uniformly integrable. In addition, we have established the result that (b) $\hat{\mu}_i - \mu_i = o_p(1)$, a.s, given *iid* sampling by SLLN. By the Continuous Mapping Theorem, $g(\hat{\mu}_i) - g(\mu_i) = o_p(1)$ Sub - indexing by i signifies that $\mu$is a function of z_i, which implies that convergence is to a random variable. By Dominated Convergence, we conclude.

□

.

Lemma 5. $\mu_i = O_p(TV(f_{x,z_i,n}, p_{x,z_i}))$. *Furthermore,*

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

A specific case of the above result is that of (Newey and Smith 2004), where $TV(f_{x,z_i,n}, p_{x,z_i}) = O_p(n^{-\xi})$ *and therefore* $\mu_i = o_p(1)$ *and if* $\frac{1}{\zeta} < \xi < \frac{1}{2}$, $\max_i \sup_\vartheta |\mu_i' m(\vartheta, x_k)| = o_p(1)$.

Proof. Consider the numerator,

$$\begin{aligned}
\frac{1}{n_s}\sum_{j=1..s} m_{j,i}(\vartheta) &= \int m_i(\vartheta) dF_{x,z_i,nS} \\
&= \int m_i(\vartheta) dP_{x,z_i} + \int m_i(\vartheta) d(F_{x,z_i,n_s} - F_{x,z_i} + F_{x,z_i} - P_{x,z_i}) \\
\text{by definition} &= \int m_i(\vartheta) d(F_{x,z_i,n_s} - F_{x,z_i} + F_{x,z_i} - P_{x,z_i})
\end{aligned}$$

Assuming that Radon-Nikodym derivatives exist with respect to the Lebesgue measure on x, we have that

$$\begin{aligned}
\int m_i(\vartheta) d(F_{x,z_i,n_s,n} - F_{x,z_i,n} + F_{x,z_i,n} - P_{x,z_i}) &= \int m_i(\vartheta)(f_{x,z_i,n_s,n} - f_{x,z_i,n}) dx + ... \\
&\quad + ... \int m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i,}) dx \\
&= o_{p,1}(1) + \int m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i}) dx
\end{aligned}$$

The second equality is using the fact that the function $m_i(\vartheta)(f_{x,z_i,n_s} - f_{x,z_i,})$ is itself dominated by $\sup_\psi m_i(\vartheta)|f_{x,z_i,n_s} - f_{x,z_i,}|$ which is integrable by $\mathbf{BD} - \mathbf{1}a$. By SLLN and Dominated convergence, we have that $\int m_i(\vartheta)(f_{x,z_i,n_s} - f_{x,z_i,}) dx \overset{a.s}{\to} 0$. Furthermore, regarding the third inequality, we make use of the implicit assumption of absolute continuity of p_{x,z_i} w.r.t $f_{x,z_i,n}$. Define a set $\mathcal{B}_{0,F} = \{x \in \mathcal{X} : F_{z_i}(x) = 0$ and $\mathcal{B}_{0,P}$ similarly. Therefore, $\mathcal{B}_{0,P} \subseteq \mathcal{B}_{0,F}$.

$$\begin{aligned}
\int m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i}) dx &= \int_{\mathcal{B}_{0,F}} m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i}) dx + \int_{\mathcal{B}_{0,F}^c} m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i}) dx \\
&= \int_{\mathcal{B}_{0,F}^c} m_i(\vartheta)(f_{x,z_i,n} - p_{x,z_i}) dx
\end{aligned}$$

It is therefore safe to assume that the events $\{x=\infty, x=-\infty\}$ belong to $\mathcal{B}_{0,P}$. Therefore,

$$\begin{aligned}\int_{\mathcal{B}^c_{0,F}} m_i(\vartheta)(f_{x,z_i,n}-p_{x,z_i})dx &\leq \sup_{(z_i,x)\in Z\times\mathcal{B}^c_{0,F}} |m_i(\vartheta)| \int_{\mathcal{B}^c_{0,F}} |f_{x,z_i,n}-p_{x,z_i}|dx \\ &\leq c\int |f_{x,z_i,n}-p_{x,z_i}|dx \leq cTV(F_{x,z_i,n},P_{x,z_i})\end{aligned}$$

Therefore, we have shown that:

$$\frac{1}{n_s}\sum_{j=1..s} m_{j,i}(\vartheta) = O_p(TV(F_{x,z_i,n},P_{x,z_i}))$$

Similarly, the denominator, given domination assumptions (3,4), we have that

$$\frac{1}{n_s}\sum_{j=1..s} m_{j,i}(\vartheta)m_{j,i}(\vartheta)' \xrightarrow{p} H > 0.$$

Also, for $\bar{M} < \infty$

$$\begin{aligned}P(\max_i \sup_\vartheta ||m(\vartheta,x_i)|| > \bar{M}n^{\frac{1}{\zeta}}) &= P(\bigcup_{k\leq n}\{\sup_\vartheta ||m(\vartheta,x_i)|| > \bar{M}n^{\frac{1}{\zeta}}\}) \\ &\leq \sum_i Pr(\sup_\vartheta ||m(\vartheta,x_i)||^\zeta > \bar{M}^\zeta n) \\ &\leq \frac{\sum_i \mathbb{E}(\sup_\vartheta ||m(\vartheta,x_i)||^\zeta \mathbf{1}(\sup_\vartheta ||m(\vartheta,x_i)||^\zeta > \bar{M}^\zeta n)}{\bar{M}^\zeta n} \\ &= \tilde{M}\mathbb{E}(\sup_\vartheta ||m(\vartheta,x_i)||^\zeta \mathbf{1}(\sup_\vartheta ||m(\vartheta,x_i)||^\zeta > \bar{M}^\zeta n) \\ &\to 0\end{aligned}$$

for $\zeta \leq 4$ due to assumption $\mathbf{BD-1}$. Therefore,

$$\max_i \sup_\vartheta ||\mu_i' m(\vartheta,x_i)|| \leq O_p(TV(F_{x,z_i,n},P_{x,z_i}))O_p(n^{\frac{1}{\zeta}}) = O_p(TV(F_{x,z_i,n},P_{x,z_i})n^{\frac{1}{\zeta}})$$

□

Lemma 6. *For any real valued function $c(x_i,\psi)$ satisfying Lemma 7 and a weighting function $\alpha_{ij} : \sum_i \alpha_{ij} = 1$ and $\max_j |\alpha_{ij}| < \infty$, then $||\frac{1}{n}\sum_i \alpha_{ij}c(x_i,\psi)|| = O_p(n^{\frac{1}{\zeta}})$*

Proof. $||\frac{1}{n}\sum \alpha_{ij}c(x_i,\psi)|| \leq \max_j |\alpha_{ij}| \max_{i,j}\sup_\psi ||c(x_i,\psi)|| = O_p(n^{\frac{1}{\zeta}})$

□

Corollary 3.2. *Any quantity of the form $|\frac{1}{n_s}\sum_j e_{i,j}\mu_i' m_j R_j|$ where $R_j$is an arbitrary x-measurable function, is $o_p(1)$ as long as $\mathbb{E}||R_j||^\zeta < \infty$ and $\zeta > 2$*

Proof. $||\frac{1}{n_s}\sum_j e_{i,j}\mu_i' m_j R_j|| \leq \max_j ||R_j|| ||\frac{1}{n_s}\sum_j e_{i,j}\mu_i' m_j|| = O_p(n^{\frac{1}{\zeta}})O_p(n^{-\frac{1}{2}}) = o_p(1)$

□

Lemma 7. ***Determination of*** μ_i**:** $\mu_i = O_p(\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta))$

Proof. The equation characterizing μ_i is $\frac{\sum_{j=1..s} m_j(\vartheta)}{\sum_{j=1..s}\kappa_j m_j(\vartheta)m_j(\vartheta)'}$ where $\kappa_{j.i} = \frac{1-e^{\mu_i' m_{j,i}(\vartheta)}}{m_{j,i}(\vartheta)'\mu_i}$. The claim is that in determining the asymptotic behaviour of $\mu_i$it is sufficient to look at the stochastic behaviour of the numerator, provided the denominator is a bounded random variable.

Case 1: $\mu_{i,0} = 0$, $w.p.1$. It has been independently shown that $\hat{\mu}_i = \mu_i + o_{p,i}(1)$ a.s. Then, $\hat{\mu}_i = o_{p,i}(1)$ a.s. Substituting in the equation of $\hat{\mu}_i$, and noticing that $\kappa_{j.i} = -1 + o_p(1)$, $\frac{\sum_{j=1..s} m_{j,i}(\vartheta)}{-\sum_{j=1..s} m_{j,i}(\vartheta)m_{j,i}(\vartheta)'+o_{p,i}(1)} = o_{p,i}(1)$. By assumption, $\sum_{j=1..s} m_{j,i}(\vartheta)m_{j,i}(\vartheta)' - o_{p,1}(1) = \mathbb{E}_{f,i}(m(\vartheta)m(\vartheta)') \neq 0$, as. If $\sum_{j=1..s} m_{j,i}(\vartheta) = \mathbb{E}_{f,i}m(\vartheta) + o_{i,p}(1)$,a.s. then it must be that $\mathbb{E}_{f,i}m(\vartheta) = o_{p,i}(1)$ a.s. Then, $P(\mu_{i,0} = 0) > P(\mathbb{E}_{f,i}m(\vartheta) \neq 0) = 1-\varepsilon$, ε arbitrarily small, which agrees with the initial assumption. Therefore, for $\mu_{i,0} = 0$ and $\mathbb{E}_{f,i}(m(\vartheta)m(\vartheta)') \neq 0$, $\sum_{j=1..s} m_{j,i}(\vartheta)$ determines $\hat{\mu}_i$.

Case 2: $\mu_{i,0} \neq 0, w.p.1$. In this case we cannot establish a direct algebraic determination of $\hat{\mu}_i$. Nevertheless, for stochastic order results, we can still establish the following result. $\hat{\mu}_i = O_p(\frac{1}{n}\sum_{j=1..s} m_j(\vartheta))$. Consider again

$$\hat{\mu}_i = \frac{\sum_{j=1..s} m_j(\vartheta)}{\sum_{j=1..s}\kappa_j m_j(\vartheta)m_j(\vartheta)'} = -\frac{\sum_{j=1..s} m_j(\vartheta)}{\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)' + \sum_{j=1..s} r(\hat{\mu}m_j(\vartheta))m_j(\vartheta)m_j(\vartheta)'}.$$

The second term in the denominator can be decomposed into

$$||\sum_{j=1..s} r(\hat{\mu}m_j(\vartheta))m_j(\vartheta)m_j(\vartheta)'|| \quad \leq \quad \begin{cases} ||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'|| & \text{if } r \in (\mu_i' m_j(\vartheta), 1) \\ ||\hat{\mu}_i||||\sum_{j=1..s}||m_j(\vartheta)||m_j(\vartheta)m_j(\vartheta)'|| & \text{if } r \in (1, \mu_i' m_j(\vartheta)) \end{cases}$$

Since we already know that $\hat{\mu}_i = \mu_i + o_{p,i}(1)$, that is $\hat{\mu}_i$ is a bounded, then $||\hat{\mu}_i||$ is also $O_p(1)$ as long as $\mu_i$exists. Using also assumption **PD-1** for P_n, it must be the case that the second term is bounded too. Let the event $r \in (1, \mu_i' m_j(\vartheta))$be signified by event E . Looking at the quantity

$$\begin{aligned} ||\mu_i|| \quad &= \quad \frac{||\sum_{j=1..s} m_j(\vartheta)||}{||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)' + \sum_{j=1..s} r(\hat{\mu}m_j(\vartheta))m_j(\vartheta)m_j(\vartheta)'||} \\ &\geq \quad \frac{||\sum_{j=1..s} m_j(\vartheta)||}{||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'(1+\mathbf{1}_{E^c})|| + ||\mu_i||||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'\mathbf{1}_{E^c}||} \end{aligned}$$

Therefore, $||\mu_i||$solves the following inequality,

$||\mu_i||||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'(2-\mathbf{1}(r \in (1, \mu_i' m_j(\vartheta)))|| + ||\mu_i||^2||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'(\mathbf{1}(r \in (1, \mu_i' m_j(\vartheta)))|| - ||\sum_{j=1..s} m_j(\vartheta)|| \geq 0$

The only admissible solution is :

$$||\mu_i|| = \begin{cases} \geq & \frac{||\sum_{j=1..s} m_j(\vartheta)||}{2||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'||} \text{ if } E^c \\ \geq & -2||\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'|| \text{ otherwise} \end{cases}$$

For some $\bar{M} < \infty$

$$\begin{aligned} Pr(||\hat{\mu}_i|| > \bar{M}||\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta)||) \quad &= \quad Pr(E^c)Pr(||\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'|| < \frac{1}{2\bar{M}}) + \ldots \\ &\quad \ldots + Pr(E)Pr(-||\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'|| \end{aligned}$$

$$\geq \frac{\bar{M}}{2}|| \sum_{j=1..s} m_j(\vartheta)||)$$

Now, the second probability of the second line is zero as $||.|| > 0$.
$Pr(||\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta)m_j(\vartheta)'|| < \frac{1}{2\bar{M}})$ can be made arbitrarily small ($< \varepsilon$), by choosing $\bar{M} = \frac{1}{2\varepsilon} >$ 0 and using $\mathbf{PD-1}$. Thus, $\mu_i = O_p(\frac{1}{n_s}\sum_{j=1..s} m_j(\vartheta))$.

□

Lemma 8. *A useful result. With regard to r products of the form:*

$$\prod_{r=1}^{R}\sum_{i=1}^{N} x_i^r$$

In the case of iid observations and for $\mathbb{E}x_i^1 x_i^2..x_i^R < \infty$

1)In the case of 4 *products,*

$$\begin{aligned}
\frac{1}{N^4}\mathbb{E}\sum_i x_i \sum_i y_i \sum_i z_i \sum_i w_i &= \frac{1}{N^4}(\begin{pmatrix} N \\ 1 \end{pmatrix} x_i y_i z_i w_i + \begin{pmatrix} N \\ 2 \end{pmatrix} (x_i y_i z_j w_j + x_i y_j z_i w_j + x_i y_{ji} z_j w_i) ... + \\
&\quad ... + \begin{pmatrix} N \\ 3 \end{pmatrix} (x_i y_j z_j w_j + x_j y_i z_j w_j + x_j y_j z_i w_j + x_j y_j z_j w_i) + ... \\
&\quad ... + \begin{pmatrix} N \\ 4 \end{pmatrix} x_i y_j z_k w_l) \\
&= O(N^{-1}) + \frac{(N-1)(N-2)(N-3)}{4!N^3}\mathbb{E}x_i \mathbb{E}x_j \mathbb{E}x_k \mathbb{E}w_l
\end{aligned}$$

2) In the case of 3 products

$$\begin{aligned}
\frac{1}{N^3}\mathbb{E}\sum_i x_i \sum_j y_j \sum_k z_k &= \frac{1}{N^3}\left[\begin{pmatrix} N \\ 1 \end{pmatrix} x_i y_i z_i + \begin{pmatrix} N \\ 2 \end{pmatrix} x_i y_j z_j + \begin{pmatrix} N \\ 2 \end{pmatrix} x_i y_i z_j + \begin{pmatrix} N \\ 3 \end{pmatrix} x_i y_j z_k\right] \\
&= O(N^{-1}) + \frac{(N-1)(N-2)}{3!N^2}\mathbb{E}x_i \mathbb{E}x_j \mathbb{E}x_k
\end{aligned}$$

2)In the case of two products,

$$\frac{1}{N^2}\mathbb{E}\sum_i x_i \sum_j y_j = \frac{1}{N^2}\left[\sum_i \mathbb{E}x_i y_i + \sum_{i \neq j}\mathbb{E}x_i y_j + \sum_{i \neq j}\mathbb{E}x_i y_i\right]$$

$$= \frac{1}{N}\mathbb{E}x_i y_i + \frac{(N-1)}{N}\mathbb{E}x_i\mathbb{E}y_j$$

3)In the case of identically distributed dependent observations, the degree of dependence will determine the summability or the rate of growth of the above covariances. For $h = i - j$, $u = k + h - i$, $w = u - h$

$$\begin{aligned}
\frac{1}{N^3}\mathbb{E}\sum_i x_i \sum_j y_j \sum_k z_k &= \frac{1}{N^3}\left[\sum_i \mathbb{E}x_i y_i z_i + \sum_{i\neq j}\mathbb{E}x_i y_j z_j + \sum_{i\neq j}\mathbb{E}x_i y_i z_j + \sum_{i\neq j\neq k}\mathbb{E}x_i y_j z_k\right] \\
&= O(N^{-2}) + \frac{2}{N^3}\sum_{i,h>0}\mathbb{E}x_i y_{i-h} z_{i-h} + \ldots \\
&\quad + \ldots \frac{2}{N^3}\sum_{i,h>0}\mathbb{E}x_i y_i z_{i-h} + \frac{4}{N^3}\sum_{\substack{i\neq h, h>0 \\ h\neq u, u>0}}\mathbb{E}x_i y_{i-h} z_{i+u-h} \\
&= O(N^{-2}) + \frac{4}{N^3}\sum_{i,h,w>0}\mathbb{E}x_i y_{i-h} z_{i+w} + \frac{4}{N^3}\sum_{i,h,w>0}\mathbb{E}x_i y_{i-h} z_{i-w}
\end{aligned}$$

4)In the case of two products, the assumption of ergodicity is sufficient.

$$\begin{aligned}
\frac{1}{N^2}\mathbb{E}\sum_i x_i \sum_j y_j &= \frac{1}{N^2}\left[\sum_i \mathbb{E}x_i y_i + \sum_{i\neq j}\mathbb{E}x_i y_j + \sum_{i\neq j}\mathbb{E}x_i y_i\right] \\
&= \frac{1}{N}\mathbb{E}x_i y_i + \frac{2}{N^2}\sum_{i,h>0}\mathbb{E}x_i y_{i-h} \\
&= \frac{1}{N}\mathbb{E}x_i y_i + \frac{2}{N}\sum_{h>0}(N-h)(\mathbb{C}ov(x_i, y_{i-h}) + \mathbb{E}x_i\mathbb{E}y_{i-h}) \\
&= \frac{1}{N}\mathbb{E}x_i y_i + \frac{2}{N}\sum_{h>0}(1-\frac{h}{N})(\mathbb{C}ov(x_i, y_{i-h}) + \mathbb{E}x_i\mathbb{E}y_{i-h})
\end{aligned}$$

In the case of non identically distributed dependent observations, different mixing assumptions can be employed to guarantee asymptotic independence such that the expectations taken are finite.

Chapter 4

Monetary Policy Rules and External Information

4.1 Introduction

What is implicitly assumed in a significant portion of the structural estimation literature is that policy makers react to the information generated by the stipulated model, whose variables are assumed as perfectly observable. In addition, while it is acknowledged that the information set of the agents of the economy is larger and different than the econometrician's information set, structural inference on monetary policy is often performed by neglecting the above fact. Despite the richness of the literature on monetary policy evaluation, the question of how external information affects structural inference on the stance of monetary policy is therefore interesting, mainly due to methodological reasons. Moreover, the imperfect measurement of concepts like the output gap, inflation, and the need for using the most model coherent interest rate data are issues that need to be addressed.

Nevertheless, the existence of different kind of measures of inflation, several types of interest rates and other measures like surveys, nowcasts and professional forecasts point towards the need to incorporate them in the process of inference. Using external information in performing estimation and inference is therefore deemed important. Beyond dealing with mis-measurement, including measures of expectations accounts for the fact that while the model is solved under rational expectations, the reality is that policy structure can change after some period and this can ultimately change the formation of expectations (Sargent 1984). An additional issue that is not addressed in this paper is the fact that not only the information set of the policy maker is rich, it is also different, in the sense that data used are ex post revised, and this makes historical analysis of monetary policy based on revised data questionable, according to (Orphanides 2003).

In order to address the question we will proceed in estimating a small standard New Keynesian model with an interest rate rule for the determination of the interest rate. While it is acknowledged that the model per se is important in terms of misspecification, complicating it would not necessarily alter our conclusions, as an even richer model would be potentially misspecified and poorly identified. The estimation of the model will be done using Bayesian methods. After estimating the model based the original information, we reestimate it using an augmented data set and observe the changes in the posterior distribution of the parameters of the interest rate rule. The way marginal distributions of parameters should change and what is expected are analyzed.

The rest of the paper is organized as follows. In Section 2 we provide a brief review of the relevant literature and in Section 3 we lay out the model used to perform the experiments. In Section 4 we describe the data and the specification of the prior distributions for the parameters. In Section 5 we discuss the implications for Bayesian inference of using external information while in Sections 6 and 7 we report the results. In section 8 we conclude.

4.2 A brief view on the related literature on methodology and monetary policy evaluation

The historical analysis of monetary policy is an important topic. Such analysis is useful for several reasons, including evaluating the effectiveness of policy and its actual interdependence with the business cycle. Many studies use a variety of time series approaches in estimating the coefficients of the Taylor rule, while others include more structure. Structure is implied by the theoretical predictions of models, i.e. conditional moments or even more by the solution to the system of equations characterizing the economy. A strand of the literature does not exploit the full set of cross equation restrictions but strives to perform inference based on a minimal set of restrictions. In the results of these studies, some authors e.g (Sargent 1999) and (Clarida, Gali, and Gertler 1999), argue that monetary policy since the 1980's has changed fundamentally e.g. focused on inflation stabilization rather than exploitation of the pre 1980's empirical fact, the Phillips. Another strand of authors e.g. (Leeper and Zha 2003) and (Canova 2006) show that policy preferences have not really changed, and that systematic monetary policy has been rather stable. Time varying approaches like (Cogley and Sargent 2001) and (Sims and Zha 2006) do not agree on the strength of the causal link between monetary policy and the rest of the economy.

Taking into account the mixed evidence, as this paper's focus is on whether external information alters substantially the inference on the stance of monetary policy, the issue of policy preference stability is not addressed and the sample is reduced accordingly.

Moreover, using bigger datasets to forecast inflation and analyze monetary policy has been largely an occupation of the 'less structural' approaches, which performed the analysis by 'exploiting' a larger information set i.e. (Stock and Watson 1999). The idea of using external data in estimating DSGE models is relatively new. For example, (Boivin and Giannoni 2006) relate this to approaches used in factor models, (Del Negro and Schorfheide 2013) use this information to improve forecasts. The mere fact that large datasets have proven to be useful in explaining the evolution of the economy and that forecasting performance is directly related to the efficient estimation of parameters is one of the reasons for the emergence of this literature. More information leads to an improvement in the efficiency of estimation, better identification of unobserved states and the disentangling of measurement error from structural shocks.

With regard to Bayesian techniques, there is a very rich literature and no attempt will be made to summarize it. Standard references are (Bauwens, Lubrano, and Richard 2000), (Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin 2013). In addition, for the Bayesian estimation of DSGE models, a standard reference is (An and Schorfheide 2007).

4.3 Model

As already stated, the simple new Keynesian model is used, which comprises of the following log-linear final form:

$$\tilde{x}_t = -\frac{1}{\sigma}(\tilde{i}_t - E_t\{\tilde{\pi}_{t+1}\}) + E_t\{\tilde{x}_{t+1}\} + e_{1t} \tag{4.1}$$

$$\tilde{\pi}_t = \beta E_t\{\tilde{\pi}_{t+1}\} + \kappa \tilde{x}_t + e_{2t} \tag{4.2}$$

$$\tilde{i}_t = \rho_i \tilde{i}_{t-1} + \phi_\pi \tilde{\pi}_{t-1} + \phi_y \tilde{y}_{t-1} + \phi_{\pi_e} E_t\{\tilde{\pi}_{t+1}\} + \phi_{y_e} E_t\{\tilde{y}_{t+1}\} + e_{3t} \tag{4.3}$$

where (1) is the Dynamic IS curve, which is the combination of the standard consumption Euler Equation, the Fisher equation and market clearing, (2) is the New Keynesian Philips curve derived from the 'Calvo' pricing problem of the production firm and (3) is the hybrid interest rate policy rule of the regulator. With regard to the shocks to the equations, e_{1t} can be considered a positive demand shock i.e. expenditure shock while e_{2t} can be a cost push shock. e_{1t} is a standard monetary policy shock. Moreover, x_t is the output gap, π_t inflation and i_t the nominal interest rate. With regard to the parameters, σ is the coefficient of relative risk aversion in the CRRA utility function, β the discount factor, ϑ is the inverse elasticity of labour supply and ξ the proportion of firms that cannot adjust prices.

Given that the above model is the prototypical New Keynesian model[1], we will not show the derivations from beginning to end (we briefly mention the basic elements of the model in the Appendix).

There are obviously weaknesses in this prototypical model, see for example (Chari, Kehoe, and McGrattan 2009). We nevertheless focus mostly on those relevant in answering our question. The specification of the interest rate rule avoids having to take care of money balances. We assume that the interest rate is set and demand for real money balances adjusts to reach equilibrium in the money market. In this sense, matters like liquidity and non standard policy measures are not relevant, and as such the recent crisis cannot be properly addressed. Including expectations in the monetary policy rule adapts to what have been long pointed out, that due to the lag of effects of monetary policy on the wider economy, the regulator needs to react also to expected inflation and output gap. With reference to the monetary policy shock, it is assumed i.i.d and any persistence in the monetary policy rule is captured by ρ_r . Moreover, the model abstracts from investment and capital accumulation. According to (Guerron-Quintana 2010), omitting variables in the estimation of a DSGE induces effects on parameters' posteriors e.g. multimodality, which is also observed in this paper.

4.4 Estimation

4.4.1 Data

Given that this is a constant coefficient DSGE model, we avoid including in the estimation pre 1980 observations and the period of after the onset of the financial crisis. The period used to estimate the model is therefore 1981Q1 to 2006Q1. We use quarterly data from the United States on Gross Domestic Product (GDP), 3-month Treasury Bill and Inflation constructed from the Consumer Price Index (CPI).[2]. With regard to filtering or detrending the data, we chose to detrend using a deterministic trend. Some series that exhibited change in trends where detrended by splitting the sample at $t*$ where $t*$ was determined by minimizing the sum of squared

[1] For the analytical derivation a standard reference is (GalÃ 2008)

[2] Data Source is the FRED database at the St Louis Fed

residuals in the corresponding subsample [3]. With regard to stochastic trends, no filtering has been done. Using a filter like Hodrick-Prescott would probably alter the power of the spectrum over business cycle frequencies and introduce spurious serial correlation in the filtered series (Canova 2007). Given that we are looking for the effects of introducing new information, using filtering would probably distort our conclusions.

Since we are concerned with the effects on the parameters' posterior distributions of adding information , changing the variability of a series will distort its true informativeness on predicting the state, something very important as illustrated later on. [4]

4.4.2 Extraneous data and the Experiment:

The experiment is done in four stages. In the first stage, the model is estimated using the aforementioned variables while in the second stage, the model is estimated by adding Inflation measures, like Producers Price Index (PPI), Personal Consumption Expenditures (PCE) and the Implicit Output Deflator (DEF). The coherence of these measures to CPI inflation is illustrated in the appendix.

In the third stage, the model is estimated using the original variables, plus measures of expectations. With regard to inflation expectations, there exist two databases, the Survey of Professional Forecasters and the U Michigan household survey. According to (Ormeno 2009), household expectations which are based on a general sense of inflation are unlikely to be a good match to CPI inflation, which is used in the model. Arguably, the Survey of professional forecasters can be a much better predictor of the state and it can also be justified with the whole methodology, that is the Rational expectations solution. Since we are imposing this, then a consistent measurement is needed. In addition, it can be said that professional forecasters have a better view of the economy, and they condition on a larger information set, which possibly nests that of households. In the second stage, we also include expected output gap measures, like manufacturing Purchasing Managers Index (PMI) which can be considered as an indicator of future productive capacity. A potential drawback of PMI is that the manufacturing sector has declined over the years, and in addition productive capacity in terms of services is hard to capture. An additional measure of expectations is SENT, the Consumer Sentiments Survey.

In the final stage, we include all of the above and other measures, e.g. U.Michigan data

[3] $t* = \arg\min_{[1,..T]}(\sum_{i=1}^{t} \tilde{\varepsilon}_i^2 + \sum_{i=t+1}^{T} \tilde{\varepsilon}_i^2)$.

[4] An alternative way to directly incorporate the trend in the estimation is to specify the observation equation in terms of first differences and a scale factor to account for the differences in the scale of the series.

(MICH) on expectations about inflation.

4.4.3 Priors

The prior distribution of parameters is considered important in performing inference, as the less information there is in the likelihood about parameters, the more effect the prior has on posterior inference. Priors should therefore reflect some objective or subjective information on the parameters. Consider the parameter vector

$$\Theta \equiv (\sigma, \beta, \vartheta, \xi, \phi_\pi, \phi_y, \rho_y, \rho_\pi, \sigma_{e1}^2, \sigma_{e2}^2, \sigma_{e3}^2)$$

and the prior distribution $g(\theta) = \prod_{i=1}^{|\Theta|} g_i(\theta_i)$ i.e. the parameters are assumed to have independent distributions. More analytically, the distributions are centered at calibrated values from other papers, summarized in (Canova 2006),: $\sigma \sim N(2, 0.75^2)$, $\beta \sim B(98, 2)$, $\vartheta \sim N(4, 1.25^2)$,$\xi \sim B(4, 2)$, $\psi_r \sim B(6, 2)$, $\phi_\pi \sim N(2.7, 0.35^2)$, $\phi_y \sim N(1, 0.15^2)$, $\rho_y \sim B(6, 2)$, $\rho_\pi \sim B(6, 2)$, $\sigma_{e1}^2 \sim \Gamma(2, 0.001)$, $\sigma_{e2}^2 \sim \Gamma(2, 0.001)$, $\sigma_{e3}^2 \sim \Gamma(2, 0.001)$, $\phi_{\pi_e} \sim N(2.7, 0.35^2)$,$\phi_{y_e} N(1, 0.15^2)$. With regard to priors for the additional parameters estimated when extraneous information is added, the following distributions are used:

$$\begin{aligned} \lambda_i &\sim N(1, 3) \\ \sigma_i &\sim \Gamma(2, 0.001) \end{aligned}$$

With regard to indeterminacy, the segment of the parameter space that generates unstable solutions to the expectational system is assigned $\log L(\theta, Y) = -\infty$. In this sense there exist restrictions imposed by the likelihood on the domain of the parameter space[5].

4.4.4 Posterior Simulation

In order to obtain draws from the posterior distribution, we use the Random Walk Metropolis Hastings algorithm with an acceptance rate between 33% and 50%. The simulation sample size, T, was increased when more data where added as there was added variance to the process and we kept every $10th$ draw after an initial burn in $\bar{L} = T - 10000$. The typical induced sample size was therefore 1000 data points.

[5]Estimation of the New Keynesian model without restricting the parameter space has been undertaken by (Thomas A. Lubik 2004) but inference is more involving as there exist observational equivalent equilibria in the determinacy and indeterminacy region.

4.5 Using external information

In this section we describe the setup of the experiments to investigate the effect on posterior inference when including extraneous information. What we essentially do is to enlarge the vector of observables in the measurement equation of the current state vector $s_t \equiv [x_t, \pi_t, i_t, e_{1t}, e_{2t}, e_{3t}]$. The new state-space system is therefore the following:

$$s_t = P(\theta)s_{t-1} + Q(\theta)v_t \tag{4.4}$$

$$d_t = \begin{bmatrix} x_t \\ \pi_t \\ i_t \\ h_t \end{bmatrix} = \begin{bmatrix} I_{\dim\Delta\times\dim\Delta} \\ \Lambda_{\dim(\tilde{h}_t)\times\dim\Delta} \end{bmatrix} \Delta s_t + u_t \tag{4.5}$$

where Δ is a selection matrix.[6]

4.5.1 Implications for Bayesian Inference

The usefulness of adding extraneous information becomes more obvious when we consider the construction of the likelihood by using the Kalman Filter. Looking at the algorithm, we see that the updated estimates of the state $s_{t|t}$ are based on the projection of the state estimation error $s_t - s_{t|t-1}$ on the prediction error $d_t - d_{t|t-1}$. A larger information set on the observables has the potential of improving the signal.

Proposition 10. *Let F_t be the filtration generated by the model at time t, and $F_{o,t} \subset F_t$ the information set generated by the observed vector $\underset{n_d\times 1}{d_t}$. Then, for an augmented vector $d_t' \equiv [d_t, \tilde{d}_t]$, where $\tilde{d}_t$ is F_t- measurable but not $F_{o,t}-$ measurable*

- $\mathbb{V}(s_{t|t}, d_t') \leq \mathbb{V}(s_{t|t}, d_t)$
- $\mathbb{V}(\hat{\theta}') \leq \mathbb{V}(\hat{\theta})$

Proof. See Appendix □

Efficiency improvements in the estimation of the latent state translate to efficiency improvements in the estimation of parameters. Given that enlarging the information set implies

[6] In our case, the original variables used in estimation are the ones allocated with the unit coefficient in the Λ matrix. In addition, we add measurement error to the variable we are adding an indicator for. If no indicator is added, then the variable is assumed to be perfectly observable. In addition, we have restricted the information gain to be solely for a single variable, but it can be easily extended to more variables.

that the likelihood function $l(\theta|Y)$ will be steeper, that is, lower $z'\Sigma_{t|t-1}z, z > 0$ for all t), this will have an effect on the posterior distribution of the parameters. Intuitively, this implies that in regions of the parameter space where the likelihood was relatively flat i.e. the posterior was not much more informative than the prior, the posterior will now be different.

Nevertheless, a change in the variance of the marginal posterior distributions does not necessarily imply uniform changes on the rest of the properties of the posterior distribution across the parameter space. As already noted, Bayesian estimates of θ i.e. $\hat{\theta} = \arg\min_{\Theta} \int \mathcal{L}(\theta,\hat{\theta})p(Y|\theta)d\theta$ where $\mathcal{L}(\theta,\hat{\theta})$ is a general loss function, are also minimizers of the Kullback Leibler Information Criterion (KLIC), $\arg\min_{\theta\in\Theta} \int log(\frac{f(y)}{p(Y|\theta)})dF(Y)$ where in this case $f(y)$ is the joint distribution of the observables. Therefore, if the model is misspecified, including more information will induce a change in the pseudo true parameters that optimize KLIC. This leads to drastic changes in the location of the marginal posterior distributions of θ, something that we will document in our empirical results. Moreover, including more information will affect the estimates of the latent states, as they are parameter driven.

What is also understood is that information will be more relevant for some states over the others, and so the efficiency gains will not be uniform over the parameter space. Given that the coefficients of the solution of the model are complicated functions of the primitives, it is difficult to see how the information gain will be allocated. With regard to identification issues, including more information can improve identification problems that are related to the 'flatness' of the likelihood; unfortunately it cannot potentially deal with population identification problems.

4.6 Results

4.6.1 Initial Results

The main object of interest are the parameters of the policy rule, and therefore inference on the deep parameters will not be emphasized. Below are the initial results of the benchmark estimation of the model, the posterior distribution of the parameters of interest.

Looking at figure 4.1, the (normalized) posterior distributions appear to be much more informative than the prior for some parameters, namely $\sigma, \vartheta, \xi, \rho_r, \sigma_m^2, \phi_{ye}, \phi_y$, something that implies that the likelihood is informative in those dimensions. On the contrary, the likelihood is comparatively less informative for $\beta, \phi_\pi, \phi_{ye}$, . Note that although the same prior probability distribution was placed on backward and forward looking components of monetary policy, a posteriori there is evidence of stronger reaction to expected inflation, as the mass of the distribution of $\phi_{\pi e}$ is shifted to the right. In addition, there is a lot of information about the persistence of monetary policy, which is concentrated around 0.65, and the variance of the monetary policy shock is much higher than where the prior placed most of the mass of the distribution.

As evident in figure 4.2, including the measure of PCE, leads to a shift in the mass of the distribution to all of the relevant variables, implying a stronger response to past inflation, lower response to expected inflation and past and expected output gap,and lower persistence to the interest rate rule. In this sense, monetary policy appears relatively less forward looking, although relative magnitudes are still high. In general, what is also noticeable is that the distributions much steeper. In addition, in some of the parameters, the support of the marginals is non overlapping with that using the original measure, implying a substantial effect on inference. The inclusion of PCE also adds relevant information to the estimates of the deep parameters, especially ξ where the effect is particularly sharp, since it is directly related to the model implied process of inflation.

With regard to figure 4.3, the Deflator is informative for all parameters; Monetary policy is stronger on the backward looking component, as the posterior for ϕ_{π} is shifted to the right while the posterior of ϕ_{π_e} shifts to the left . The effect on ϕ_{ye} is also significant as the two posteriors do not overlap. Policy appears to be much less responsive to the output gap. In addition, the interest rate rule is much less persistent and the variance of the monetary policy shock smaller. As it regards the deep parameters, the same effect on ξ is observed;β and θ are also sharply concentrated.

In the next section we report results from experiments using information on agent expectations.

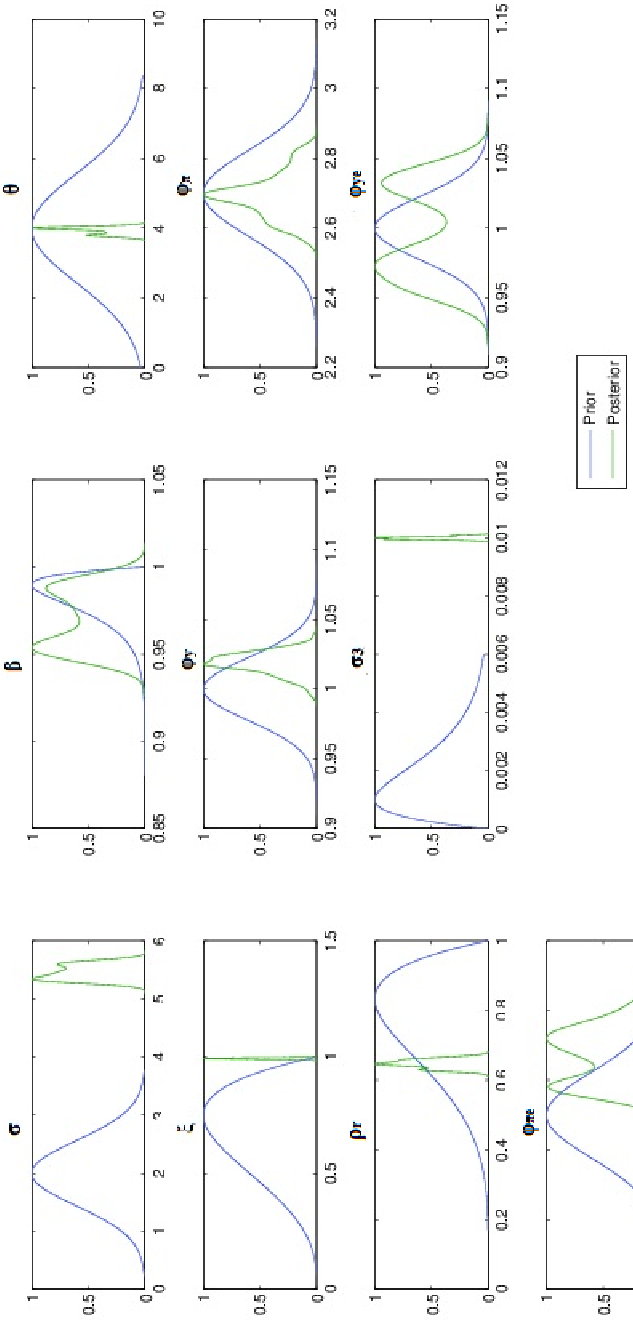


Figure 4.1: Comparison of prior and posterior

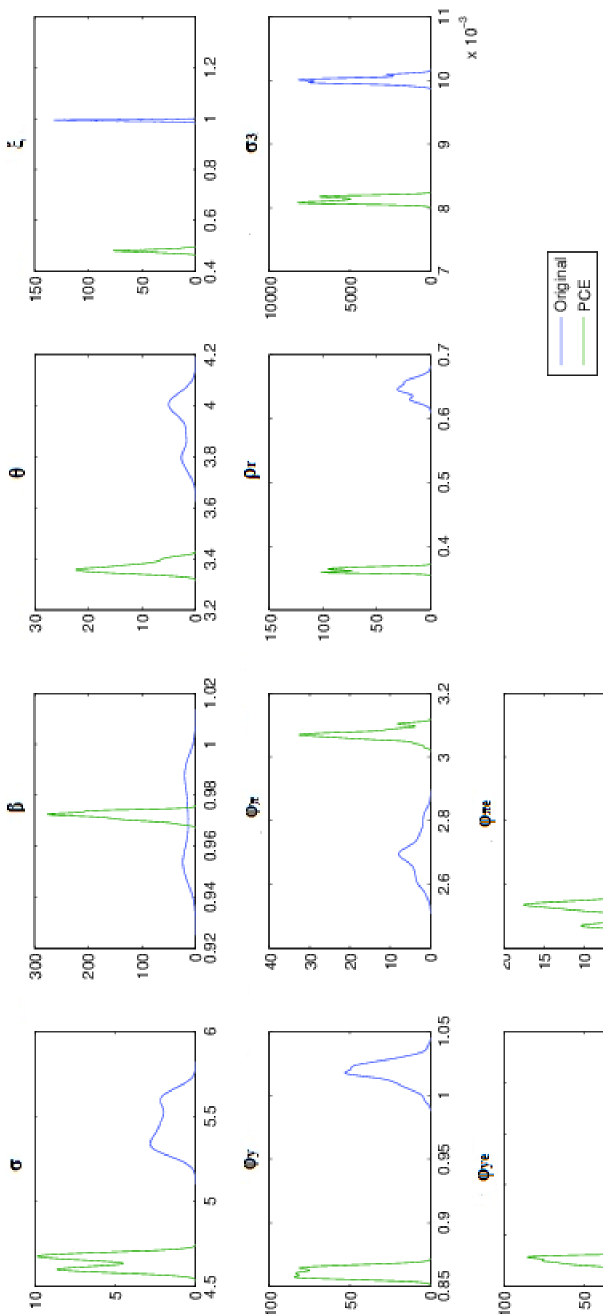


Figure 4.2: Comparison of Posterior with CPI and Posterior with CPI and PCE

4.7 Results on noisy measures:

4.7.1 Surveys and Expectations

4.7.1.1 Inflation expectations and different information sets

Incorporating expectations involves using the state vector such that to link the observed data on expectations to the model's latent variables. For example, for $\tilde{\pi}^e_{t+1}$ denoting the observed inflation expectations, then the measurement equation becomes as follows:

$$d_{t,\pi} = \begin{bmatrix} \pi_t \\ \tilde{\pi}^e_{t+1} \end{bmatrix} = \Lambda_\pi \begin{bmatrix} I_{dim(s_t)} \\ P(\theta)_\pi \end{bmatrix} s_t + u_t \tag{4.6}$$

since $\tilde{\pi}^e_{t+1} = E_t\{\pi_{t+1}\} + u_t = E_t\{P(\theta)_\pi s_t + Q(\theta)v_{t+1}\} + u_t = P(\theta)_\pi s_t + u_t$.

A slight complication arises because the 'inflation expectations' quarterly data are for the 1-year horizon while the model involves expectations over the next quarter. Nevertheless, we can still use the structure implied by the DSGE solution to map the data to the model expectations, that is $\tilde{\pi}^e_{t+k} = E_t\{\pi_{t+k}\} + u_t = [P(\theta)^k]_\pi s_t + u_t$.

As mentioned before, using measures of expectations involves exploiting the fact that, under rational expectations, the information set of the econometrician $\mathcal{O}_t$, the information set of the professional forecasters,$\mathcal{H}^p_t$, and that of the agents of the economy $\mathcal{H}^a$, are assumed to be related as such: $\mathcal{O}_t \subset \mathcal{H}^p_t \subset \mathcal{H}^a_t$. In terms of conditional expectations, we can see that if we assume that $\mathcal{H}^p_t = \mathcal{O}_t + J_t$, $\mathcal{O}_t \perp J_t$, then $E(d_{t+1}|\mathcal{H}^p_t) = E(d_{t+1}|\mathcal{O}_t) + E(d_{t+1}|J_t) = E(d_{t+1}|\mathcal{O}_t) + u_t$, something that rationalizes the operationalization of how to include inflation expectations in our estimation.[7]

Again, including expectations measures in the observables implies that the forecast error e_t will also involve information useful for updating the forecast of the state. Nevertheless, since the link of observed and actual expectations is model dependent, model misspecification can induce further uncertainty and therefore signal extraction is also affected.

4.7.1.2 Results with Expectations measures

In this case, including measures of expected inflation, leads to more information about the relevant parameter, $\phi_{\pi e}$, but relatively less for ϕ_π. Monetary policy seems less responsive to expected inflation. In addition, no significant information is added to other parameters like ξ.

In this case, PMI is not directly related in the sense of measurement to the output gap[8];

[7] $\mathcal{O}_t + J_t$ corresponds to the set $\{o_t + j_t : o_t \in \mathcal{O}_t \wedge j_t \in J_t\}$

[8] The expectations are again linked to the state variables of the DSGE through $P(\theta)^k_y$, using the same k as inflation

nevertheless, it shifts a significant amount of the mass of ϕ_y , ϕ_{ye} to the left and right respectively. Monetary policy appears more responsive to expected output gap; more that what we would infer using the original dataset. In addition, policy is again less persistent and the variance of the monetary policy shock lower.

The consumer sentiments index appears to be equally informative on the policy rule coefficients on inflation and the output gap; again, there is relatively more evidence of lagging rather than foreword looking monetary policy in terms of inflation. What is also noticeable is the fact that adding consumer sentiments carries a lot of information about σ, the coefficient of relative risk aversion and ϑ, the inverse elasticity of labour supply.

4.7.1.3 Including all measures

The inclusion of all measures has let to much more informative posteriors, as evident in Figure 4.7. All in all, the addition of extra information has affected the posteriors of the parameters of the monetary policy rule in the following way: ϕ_π is evidently higher and $\phi_{\pi e}$ lower. Monetary policy appears less responsive to lagged output gap. The posterior of ϕ_{ye} is also much more concentrated than the original. Monetary policy is much less persistent than originally inferred; what this implies is that effective monetary policy has placed more weight in the systematic component rather than keeping the previous interest rate level. In addition, the variance of the monetary policy shock is also lower; some of the variance is attributed to measurement error but we cannot disentangle which variable as roughly the same change in the posterior was observed for all experiments. The posterior distribution of ξ is much more reasonable than with the original information, which implied that more than 10% firms would not have changed their prices after 5 years! With regard to β, the posterior is concentrated over values that imply relatively high equilibrium interest rates, 6-8%. The Frisch elasticity is also higher than previously inferred.

---

expectations. It can be argued though that choosing the same k is restrictive. Same approach is followed for consumer expectations (SENT).

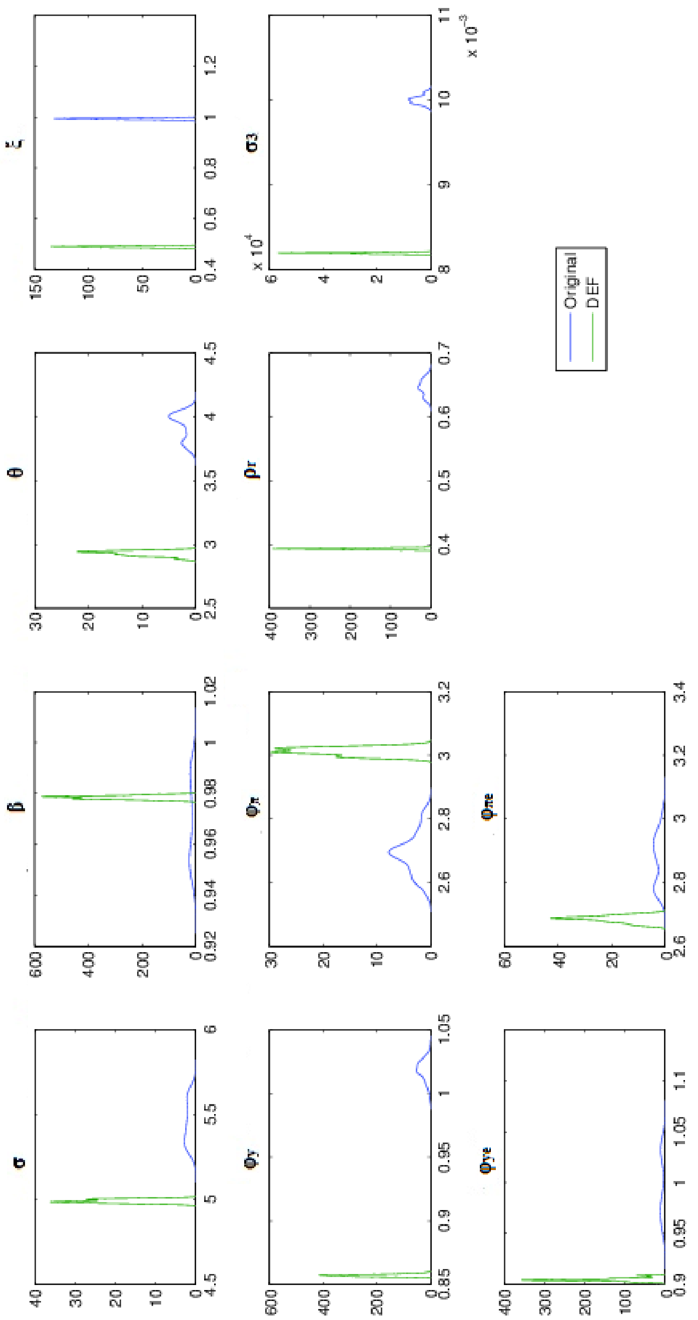


Figure 4.3: Comparison of Posterior with CPI and Posterior with CPI and DEF

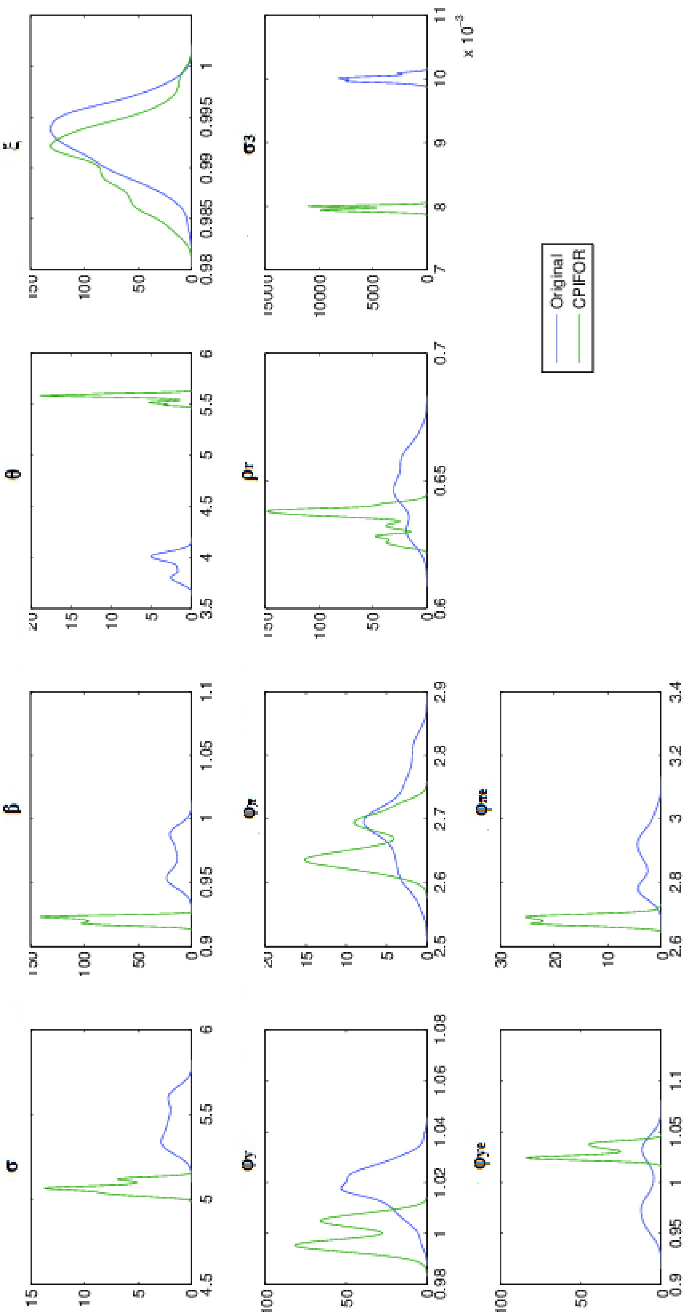


Figure 4.4: Comparison of original Posterior and Posterior with CPI forecast

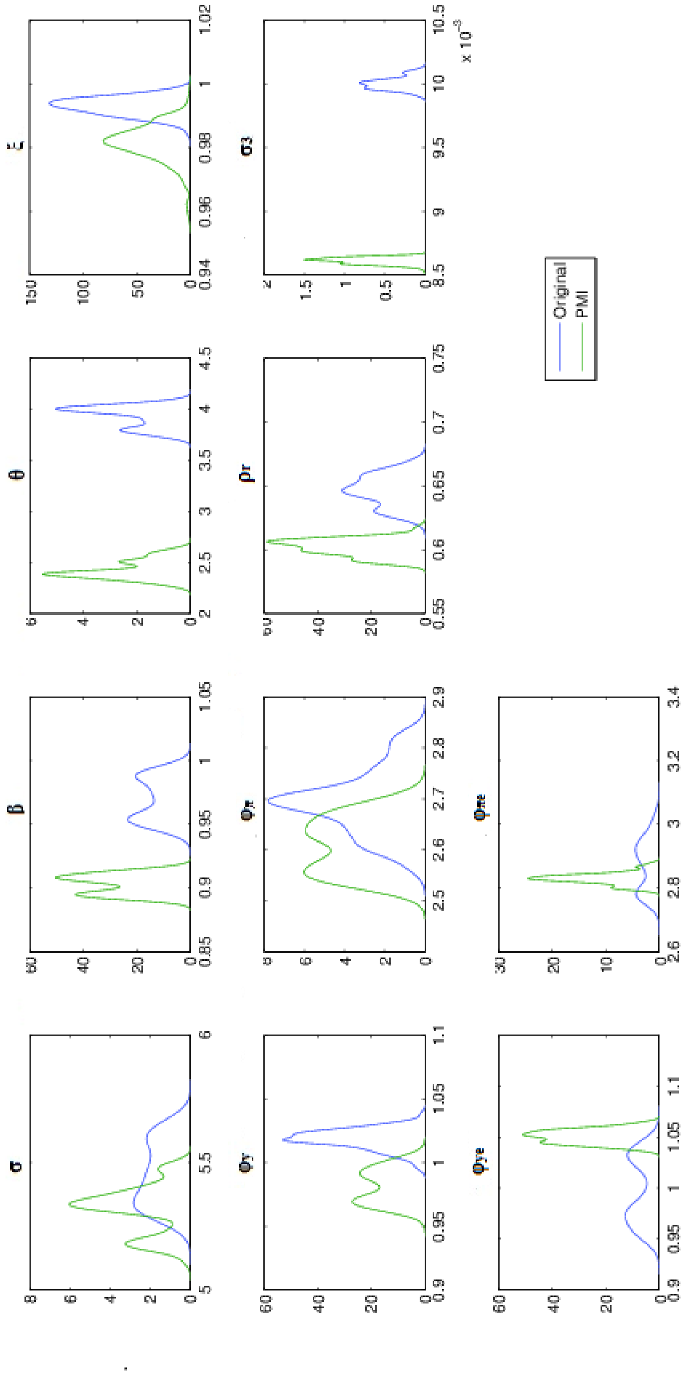


Figure 4.5: Comparison of original Posterior and Posterior with PMI

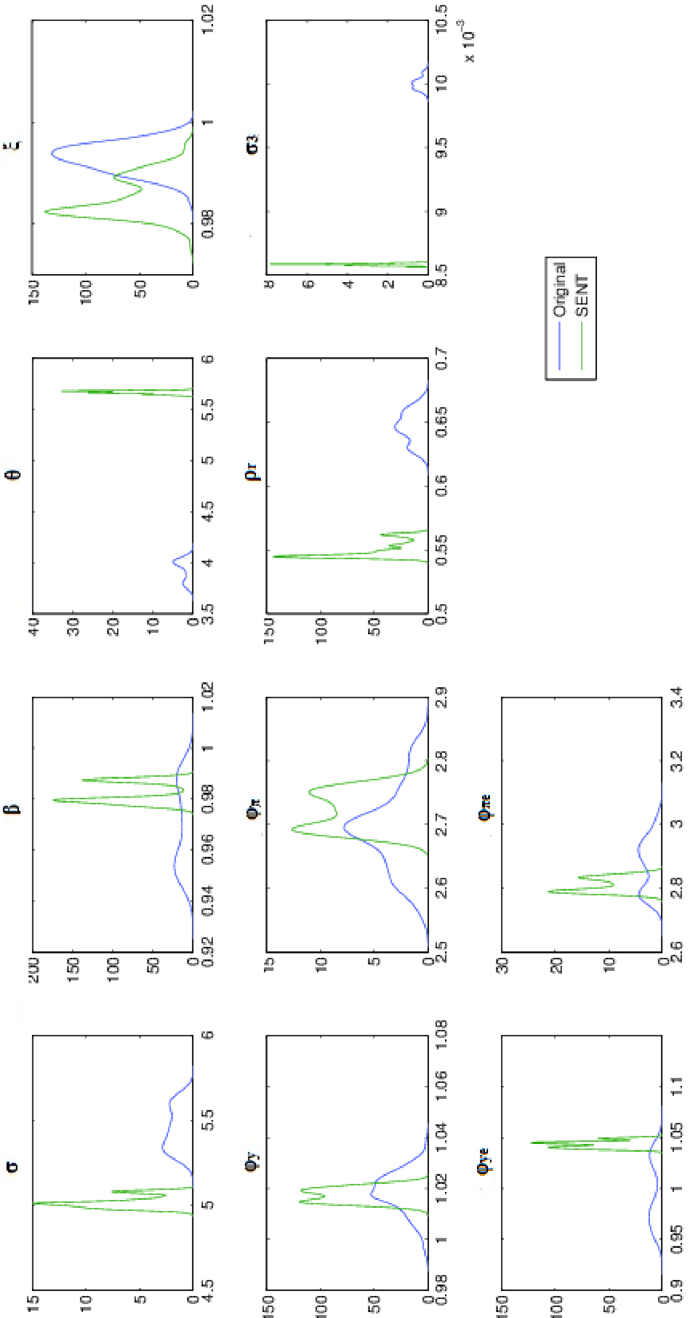


Figure 4.6: Comparison of original Posterior and Posterior with SENT

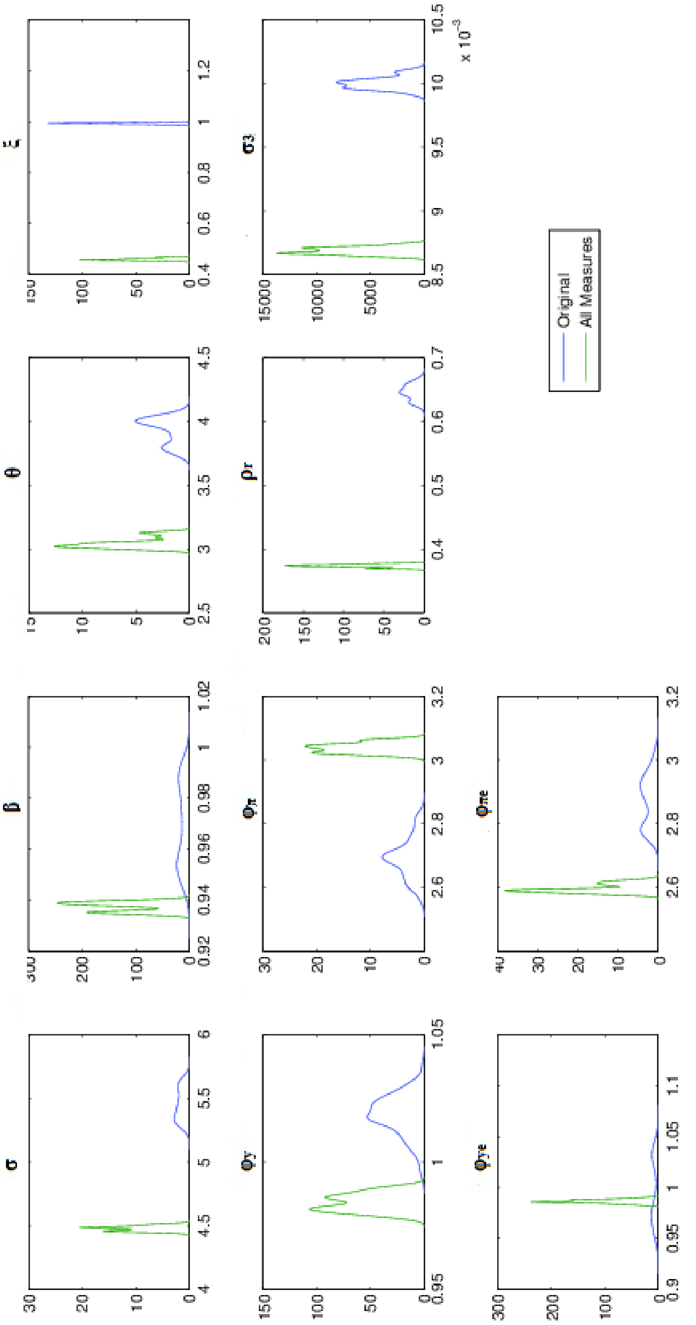


Figure 4.7: The measures included are: CPI, CPIFOR, DEF, DEFFOR, CPE, PPI, MICH, PMI, SENT

4.7.2 Marginal Likelihood Comparison

The - joint and marginal - posterior distribution of the parameters is the main workhorse of inference in Bayesian estimation. In order to be able to compare the different specifications of the observables, we can use the marginal likelihood, $p(y_t|M_i)$ which in principle can be obtained by multidimensional integration over the parameter dimensions.

Nevertheless, we proceed in computing the marginal likelihood following (Geweke 1999), that is using the modified harmonic mean estimator which is based on the identity $\frac{1}{p(y)} = \int \frac{f(\theta)}{L(\theta|Y)g(\theta)} p(\theta|Y)d\theta$ where $\int f(\theta)d\theta = 1$. Geweke (1999) proposed $f(\theta)$ to be the truncated multivariate $N(\bar{\theta}, V(\theta))$, where the arguments are the posterior mean and variance, and the truncation is defined by $\mathbf{1}\{(\theta-\bar{\theta})V_\theta^{-1}(\theta-\bar{\theta}) \leq F_{\chi^2}^{-1}(\tau)\}$ for $\tau \in (0,1)$. So the harmonic mean estimator is:

$$\hat{P}_{m_j}(Y) = \left[\frac{1}{n_s} \sum_{i=1}^{n_s} \frac{f(\theta^i)}{L(\theta^i|Y)g(\theta^i)} \right]^{-1}$$

The marginal log likelihood of the NK model using the initial information set and $\tau = 0.9$, is -1731.9014. When all extraneous information is added, the marginal log likelihood of the model deteriorates to -3125.741.

Although this observation initially seems contradictory, we can claim that while marginal likelihood is computed by evaluating the likelihood with draws from the prior distribution, $p_{m_j}(Y) = \int p(Y|\theta)p(\theta)d\theta$, extraneous information is informative for the parameters only a posteriori. In fact, by augmenting with more observables the same model m_j, it is inevitable that the in sample fit to the augmented data set will deteriorate. What is really exploited by augmenting the estimation with more data is the forecasting error of the observables. With regard to prediction, to the contrary, more information improves forecasts as evident from the predictive density, $p(Y_{T+1:T+h}|Y_{1:T}) = \int p(Y_{T+1:T+h}|Y_{1:T},\theta)p(\theta|Y_{1:T})d\theta$. Forecasts are made with respect to the posterior distribution of the parameters, which is indeed more informative than the prior, and even more informative with the extra information.

4.8 Summary and Conclusion

This paper has explored the dimensions over which additional information can help in improving inference when analyzing monetary policy through a DSGE model. Additional information includes several measures of inflation and the productive potential of the economy, surveys on inflation expectations and consumer confidence, and professional forecasts. The general result is that augmenting the model with more observables leads to a substantial change in the posterior distributions not only of the monetary policy rule parameters, but also the deep parameters. External information has both a statistical and economic interpretation within the discipline of rational expectations, and is indeed crucial in performing better inference on historical monetary policy structure. Further exercises could look at whether external information is useful in improving the forecasting performance for key macroeconomic variables. Improving the estimates of unobserved states is likely to be the main driver of possible forecast improvements. Another avenue for future research is to allow for indeterminacy. Since under indeterminacy belief shocks (sunspots) are also additional unobserved state variables, using external information can possibly improve their identification and measurement.

4.9 Appendix

4.9.1 The New Keynesian model in a nutshell

The representative consumer maximizes over an infinite horizon expected utility, where instantaneous utility belongs to the CRRA class,

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{(1-l_t)^{1-\vartheta}}{1-\vartheta} \right) \tag{4.7}$$

which is additive in consumption and labour supply. Consumption is aggregated by $C_t = \left(\int_{(0,1)} C(i)_t^{1-\varepsilon} di \right)^{\frac{\varepsilon}{1-\varepsilon}}$, where ε is the elasticity of substitution between goods. Maximization is subject to a sequence of budget constraints:

$$\int_{(0,1)} C(i)_t P_t(i) di + Q_t B_t \leq B_{t-1} + w_t N_t$$

The consumption -savings decision leads to the typical Euler equation,

$$Q_t = \beta E_t \{ \frac{U_{c,t+1} P_t}{U_{c,t} P_{t+1}} \}$$

where $P_t = \left(\int_{(0,1)} P(i)_t^{1-\varepsilon} \right)^{\frac{\varepsilon}{1-\varepsilon}}$ is the aggregate price, and the allocation of total consumption to different goods leads to the following demand function,

$$C_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\varepsilon} C_t \tag{4.8}$$

The typical intratemporal tradeoff between consumption and leisure is:

$$-\frac{C_t^\sigma}{N_t^\vartheta} = \frac{w_t}{P_t}$$

As it regards the production sector, which is monopolistically competitive, the proportion of $1-\xi$ firms that get to re-optimize, choose prices P_t^* to maximize profits,

$$E_t \sum_{j=0}^{\infty} \xi^j Q_{t,t+j} \left(P_t^* Y_{t,t+j} - \tau_{t+j}(Y_{t,t+j})\right)$$

with a constant returns to scale labour intensive production function, $Y_t(i) = Z_t N_t(i)$ and cost function $\tau_t(Y_t)$, subject to the the isoelastic consumer demand function (4.8) . The resulting first order condition is

$$E_t \sum_{j=0}^{\infty} \xi^j Q_{t,t+j} \left(P_t^* Y_{t,t+j} - mc_{t+j}(Y_{t,t+j})\right) = 0 \tag{4.9}$$

Equilibrium $Y_t = C_t$ and further manipulations lead to the New Keynesian Phillips Curve outlined above.

Proof. of Proposition 10

1. For a general state space model, defined by

$$\begin{aligned} s_t &= P s_{t-1} + Q\varepsilon_t \\ y_t &= R s_t + u_t \end{aligned}$$

Recall that from the Kalman Filter recursions, for $s_{t|t-1} \equiv \mathbb{E}(s_t|\mathcal{F}_{o,t-1})$ and $d_{t|t-1} \equiv \mathbb{E}(d_t \mathcal{F}_{o,t-1})$

$$s_{t|t} = s_{t|t-1} + \mathcal{K}_t g_t$$

where $\tilde{\mathcal{K}}_t \equiv \Omega_{t,t-1} R \Sigma_{t,t-1}^{-1}$ is the orthogonal projection of $s_t - s_{t|t-1}$ on $d_t - d_{t|t-1} \equiv g_t$, the prediction error on time t.

Therefore, conditional on $\mathcal{F}_{o,t-1}$ (or any $\mathcal{F}_{t-1}$ measurable set)

$$\begin{aligned} \mathbb{V}(s_{t|t}) &= \tilde{\mathcal{K}} \Sigma_{t,t-1} \tilde{\mathcal{K}}' \\ &= \Omega_{t,t-1} R \Sigma_{t,t-1}^{-1} R' \Omega_{t,t-1} \end{aligned}$$

Augmenting d_t to d'_t, implies that $\Sigma_{t,t-1} \geq \Sigma'_{t,t-1}$ as the dimension of observables increases. Therefore, $\mathbb{V}(s_{t|t,d'_t}) \leq \mathbb{V}(s_{t|t,d'_t})$. Another way of showing this is to set $d'_t \equiv (d_t \mathcal{Q}_{d'_t}, d_t \mathcal{Q}^{\perp}_{d'_t}) \equiv (\hat{d}_t, z_t)$ where z_t is orthogonal to d_t by construction as $\mathcal{Q}_{d'_t}$ is a projection matrix. Then,

$$\mathbb{V}(s_{t|t,d'_t}) \equiv \mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g'_t))$$

where $g'_t \equiv d'_t - d_{t|t-1}$ By the law of total variance,

$$\mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t)) = \mathbb{E}(\mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t)|g'_t)) + \mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t, g'_t))$$

and since $\mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t))$is positive for almost all $g'_t \in \mathcal{F}_t \setminus \mathcal{F}_{o,t}$,

$$\begin{aligned}\mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t)) &\geq \mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t, g'_t)) \\ &= \mathbb{V}(\mathbb{E}(s_t - s_{t|t-1}|g_t, z_t))\end{aligned}$$

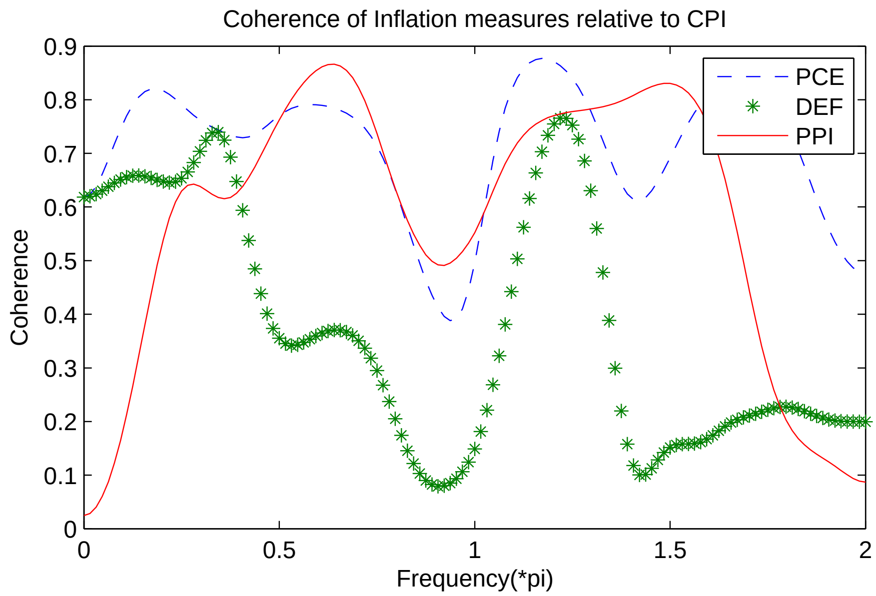
2. We have shown that $\mathbb{V}(s_{t|t,d'_t}) \leq \mathbb{V}(s_{t|t,d_t})$ conditional on $\mathcal{F}_{o,t}$ for any $t \leq T$ and conditional on any $\mathcal{F}_t$ measurable set. This implies that $\mathbb{V}(d'_{i,t+1|t}) = P\mathbb{V}(s_{t|t,d'_t})P^T$ for each variable d_i. We use the fact that the posterior density $\mathbb{P}(\theta|d_{t=1:T})$ converges in total variation distance to $N(\Delta_{n,\theta_0}, I^{-1}_{\theta_0})$ [9] where Δ_{n,θ_0} is the random score and $I^{-1}_{\theta_0}$ is the inverse of the information matrix. I_{θ_0} is inversely related to the Hessian of the log likelihood function, $-\mathbb{E}\frac{\partial^2 log(\mathbb{P}(\theta|d_{t=1:T}))}{\partial\theta\theta^T}$, and in turn the Hessian is inversely related to $\Sigma^\star = R\Omega^\star R^T + \Sigma_u$ with $\Omega^\star \equiv \lim_{t\to\infty} \Omega_{t|t-1} = P\Omega^\star P' + Q\Sigma_\varepsilon Q^T$. Since the dimension of ε_t is independent of the dimension of d_t and $\Omega^\star(d') < \Omega^\star(d)$, $\Sigma^\star(d') < \Sigma^\star(d)$ and $I_{\theta_0}(d') > I_{\theta_0}(d)$.

□

[9] Bernstein-von Mises Theorem, see (der Vaart 2000)

4.9.2 Coherence of other inflation measures to CPI inflation:

With regard to the measures of inflation used in the estimation, below is plotted their coherence to CPI inflation, $C_{x,CPI} = \frac{|P_{x,CPI}|^2}{P_{x,CPI}P_{y,CPI}}$, which is a frequency domain measure of cross correlation and under assumptions, of least squares predictability, something directly relevant to the methodology followed. As is evident, none of them is perfect over any frequency. At business cycle frequencies, that is for quarterly data at the range of $[\frac{\pi}{6}, \frac{\pi}{10}]$ their coherence lies between 0.4 and 0.8. PCE coherence is much more stable over all frequencies, which makes sense as by definition its closer to CPI inflation.



Bibliography

ACKERBERG, D., J. GEWEKE, AND J. HAHN (2009): “Comments on "Convergence Properties of the Likelihood of Computed Dynamic Models,” *Econometrica*, 77(6), 2009–2017.

ALMEIDA, C., AND R. GARCIA (2014): “Economic Implications of Nonlinear Pricing Kernels,” Discussion paper, AFA 2009 San Francisco Meetings Papers.

AN, S., AND F. SCHORFHEIDE (2007): “Bayesian analysis of DSGE models,” *Econometrics Review*, 26, 187–192.

ANDREWS, D., AND X. SHI (2013): “Inference for Parameters Defined by Conditional Moment Inequalities,” *Econometrica*, 81(2), 609–666.

ANDREWS, D. W. (1993): “An introduction to econometric applications of empirical process theory for dependent random variables,” *Econometric Reviews*, 12(2), 183–216.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (2000): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

BENIGNO, P. (2004): “Optimal monetary policy in a currency area,” *Journal of International Economics*, 63(2), 293–320.

BILLINGSLEY, P. (1961): “The Lindeberg-Levy 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.

BOIVIN, J., AND M. GIANNONI (2006): “DSGE Models in a Data-Rich Environment,” Working Paper 12772, National Bureau of Economic Research.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2012): "Set Identified Linear Models," *Econometrica*, 80(3), 1129–1155.

BUERA, F. J., AND B. MOLL (2015): "Aggregate Implications of a Credit Crunch: The Importance of Heterogeneity," *American Economic Journal: Macroeconomics*, 7(3), 1–42.

CANOVA, F. (2006): "Monetary Policy and the Evolution of the US Economy," CEPR Discussion Papers 5467, C.E.P.R. Discussion Papers.

CANOVA, F. (2007): *Methods for Applied Macroeconomic Research*, no. v. 13 in Methods for applied macroeconomic research. Princeton University Press.

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.

CHARI, V. V., P. J. KEHOE, AND E. R. MCGRATTAN (2000): "Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?," *Econometrica*, 68(5), pp. 1151–1179.

——— (2007): "Business Cycle Accounting," *Econometrica*, 75(3), 781–836.

——— (2009): "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1(1), 242–66.

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.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models1," *Econometrica*, 75(5), 1243–1284.

CHERNOZHUKOV, V., E. KOCATULUM, AND K. MENZEL (2012): "Inference on Sets in Finance," .

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

CHETTY, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80(3), 969–1018.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113(1), 1–45.

CLARIDA, R., J. GALI, AND M. GERTLER (1999): "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37(4), 1661–1707.

COGLEY, T., AND T. SARGENT (2001): "Evolving Post-World War II U.S. Inflation Dynamics," Working Papers 2132872, Department of Economics, W. P. Carey School of Business, Arizona State University.

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.

DEL NEGRO, M., AND F. SCHORFHEIDE (2013): *DSGE Model-Based Forecasting*vol. 2, Part A of *Handbook of Economic Forecasting*, chap. 2, pp. 57 – 140. Elsevier.

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.

DER VAART, A. V. (2000): *Asymptotic Statistics*. Cambridge University Press.

DOUKHAN, P., P. MASSART, AND E. RIO (1995): "Invariance principles for absolutely regular empirical processes," *Annales de l'institut Henri PoincarÃ© (B) ProbabilitÃ©s et Statistiques*, 31(2), 393–427.

GALÃ, J. (2008): *Introduction to Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework [Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Ke*Introductory Chapters. Princeton University Press.

GALI, J., AND T. MONACELLI (2005): "Monetary Policy and Exchange Rate Volatility in a Small Open Economy," *The Review of Economic Studies*, 72(3), 707–734.

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.

GELMAN, A., J. CARLIN, H. STERN, D. DUNSON, A. VEHTARI, AND D. RUBIN (2013): *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

GEWEKE, J. (1999): "Using simulation methods for Bayesian econometric models: Inference, development and communication," *Econometric Review*, pp. 1–126.

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.

GUERRON-QUINTANA, P. A. (2010): "What you match does matter: the effects of data on DSGE estimation," *Journal of Applied Econometrics*, 25(5), 774–804.

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

HANSEN, L. P. (2013): "Uncertainty Outside and Inside Economic Models," (2013).

HANSEN, L. P., AND T. J. SARGENT (2005): "Robust estimation and control under commitment," *Journal of Economic Theory*, 124(2), 258–301.

——— (2013): *Recursive Models of Dynamic Linear Economies*, The Gorman Lectures. Princeton University Press.

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.

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 S. NG (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79(6), 1995–2032.

KOMUNJER, I., AND G. RAGUSA (2016): "Existence and characterization of conditional density projections," *Econometric Theory*, pp. 1–41.

LEEPER, E. M., AND T. ZHA (2003): "Modest policy interventions," *Journal of Monetary Economics*, 50(8), 1673 – 1700.

LIAO, Y., AND W. JIANG (2010): "Bayesian analysis in moment inequality models," *The Annals of Statistics*, 38(1), 275–316.

LUTTMER, E. G. J. (1996): "Asset Pricing in Economies with Frictions," *Econometrica*, 64(6), pp. 1439–1467.

MARIMON, R., AND A. SCOTT (eds.) (1998): *Computational Methods for the Study of Dynamic Economies*. Oxford University Press.

MAVROEIDIS, S., M. PLAGBORG-MOLLER, AND J. H. STOCK (2014): "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," *Journal of Economic Literature*, 52(1), 124–88.

MOON, H. R., AND F. SCHORFHEIDE (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80(2), 755–782.

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.

ORMENO, A. (2009): "Inflation expectations, learning, and DSGE model estimation," *Mimeo UPF*.

ORPHANIDES, A. (2003): "Monetary policy evaluation with noisy information," *Journal of Monetary Economics*, 50(3), 605–631.

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

POLITIS, D. N., AND J. P. ROMANO (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89(428), 303 – 1313.

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

SARGENT, T. (1999): *The Conquest of American Inflation*, Princeton paperbacks. Princeton University Press.

SARGENT, T. J. (1984): "Autoregressions, Expectations, and Advice," *American Economic Review*, 74(2), 408–15.

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

SCHORFHEIDE, F., AND K. I. WOLPIN (2016): "To Hold Out or Not to Hold Out," Discussion paper.

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

SIMS, C. A., AND T. ZHA (2006): "Were There Regime Switches in U.S. Monetary Policy?," *American Economic Review*, 96(1), 54–81.

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

STOCK, J. H., AND M. W. WATSON (1999): "Forecasting inflation," *Journal of Monetary Economics*, 44(2), 293–335.

SU, C.-L., AND K. L. JUDD (2012): "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80(5), 2213–2230.

THOMAS A. LUBIK, F. S. (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *The American Economic Review*, 94(1), 190–217.

VAN DER VAART, A. (1998): *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

WANG, P., AND Y. WEN (2012): "Hayashi meets Kiyotaki and Moore : A theory of capital adjustment costs," *Review of Economic Dynamics*, 15(2), 207–225.

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

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52(1), 143–162.