Efficient Estimation with Latent Variables

by

Michael Bedard
B. S., University of New Hampshire, 2008
M. A., University of New Hampshire, 2012
Sci. M., Brown University, 2017

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Economics at Brown University

Providence, Rhode Island
May 2019
This dissertation by Michael Bedard is accepted in its present form by
the Department of Economics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date ______________

Eric Renault, Director

Recommended to the Graduate Council

Date ______________

Susanne Schennach, Reader

Date ______________

Adam McCloskey, Reader

Approved by the Graduate Council

Date ______________

Andrew G. Campbell
Dean of the Graduate School
Vita

Acknowledgements

I would like to thank Eric Renault, Susanne Schennach and Adam McCloskey for their support and encouragement during my time at Brown. I am also very grateful to Emily Oster for her support as placement director, and to Angelica Spertini for all her work at the Economics Department. All the effort in this dissertation is dedicated to my wife Leslie.
# Contents

List of Tables viii  

List of Figures ix  

1 Efficiency Bounds with Latent Variables 3  

1.1 Introduction 3  
1.2 Examples 5  
1.2.1 Measurement error 5  
1.2.2 Linear factor analysis 9  
1.2.3 Multivariate volatility model 14  
1.3 Efficiency in moment restriction models 14  
1.3.1 Intuition 15  
1.3.2 Preliminaries 19  
1.3.3 Nested parametric models 24  
1.3.4 Fisher information and efficiency 28  
1.3.5 Feasible efficiency bounds 29  
1.4 Conclusion 31  
1.5 Appendix: Proofs 31  
1.5.1 Proofs of Propositions 31  
1.5.2 Proofs of Lemmata 38  
1.5.3 Proofs of Theorems 47  

2 Estimation with Latent Variables 48  

2.1 Introduction 48
2.2 Two-step estimation in the unconditional model
   2.2.1 Preliminaries
   2.2.2 Consistency
   2.2.3 Asymptotic Normality
2.3 Monte Carlo results
2.4 Empirical example: long-run risk model
2.5 One-step efficient estimator for unconditional moment restrictions
2.6 One-step estimator for conditional moment restrictions
2.7 Conclusion
2.8 Appendix: Proofs
   2.8.1 Proofs of Propositions
   2.8.2 Proofs of Lemmata
   2.8.3 Proofs of Theorems
2.9 Appendix: Figures and Tables

3 Efficient Errors-in-Variables with Multiple Proxies
   3.1 Introduction
   3.2 Model and Estimation
      3.2.1 Instrument Variables Estimation
      3.2.2 Andersson and Møen (2016) Estimation
      3.2.3 Chalfin and McCrary (2017) Estimation
      3.2.4 ELVIS Estimation
   3.3 Example: bivariate model
      3.3.1 Monte Carlos Results
   3.4 General Case
   3.5 Conclusion
   3.6 Appendix: Proofs
   3.7 Appendix: Figures and Tables
List of Tables

2.1 Monte Carlo results for the two-step efficient ELVIS estimator and the minimum
distance estimator of Gospodinov et al. (2017). ................................. 90

2.2 Estimation Results for the two-step efficient ELVIS estimator, the minimum distance
estimator of Gospodinov et al. (2017) and the indirect inference estimator of Grammig
and Küchlin (2018). ................................................................. 90

3.1 Monte Carlo results for the two-step efficient ELVIS estimator, the Chalfin McCrary
estimator, and the Andersson Møen estimator. ................................. 108
List of Figures

2.1 Monte Carlo sampling distributions for the two-step efficient ELVIS estimator and the minimum distance estimator of Gospodinov et al. (2017) with $\beta = 1, 1/4$. . . . . 88

2.2 Monte Carlo sampling distributions for the two-step efficient ELVIS estimator and the minimum distance estimator of Gospodinov et al. (2017) with $\beta = 1/8, 0$. . . . . 89

3.1 Monte Carlo sampling distributions for the two-step efficient ELVIS estimator, the Chalfin McCrary estimator, and the Andersson Møen estimator. Sample size n=200. 108

3.2 Monte Carlo sampling distributions for the two-step efficient ELVIS estimator, the Chalfin McCrary estimator, and the Andersson Møen estimator. Sample size n=500. 109
Introduction

This dissertation studies the efficient estimation of semiparametric models that contain latent variables. The first two chapters develop a framework for efficient estimation of moment restriction models containing latent variables addressing the relevant efficiency bounds and estimation procedures respectively. The third chapter applies this framework to the classical errors-in-variables problem when multiple proxies are available and analyze the efficiency properties of existing estimators.

In Chapter 1, I derive semiparametric efficiency bounds for moment restriction models that contain latent variables. To derive bounds on asymptotic efficiency, I exhibit a parametric model nested within the moment model in which the inverse of the Fisher information can be seen as an efficiency bound. This approach to semiparametric efficiency extends the argument that Back and Brown (1992) used to give a simple proof of the semiparametric efficiency of the generalized method of moments estimator. The parametric model I construct may be considered least favorable for the estimation of θ.

In Chapter 2, I suggest some procedures for efficient estimation of the models considered in the first chapter. My approach to estimation of these models is to adapt the approach of Schennach (2014) for accommodating latent variables in partially identified moment restriction models. Schennach’s approach, called entropic latent variable integration via simulation (ELVIS), eliminates latent variables from the moment function by integrating against a cleverly chosen conditional probability distribution. After the latent variables are eliminated these models can be estimated using a generalized method of moments (GMM) or generalized empirical likelihood (GEL) type estimator.

Chapter 3 is joint work with Eric Renault. We consider the efficient estimation of errors-in-variables models when multiple proxies are available. By applying the results of the first two chapters we derive sufficient conditions for the estimators introduced by Andersson and Møen (2016) and
Chalfin and McCrary (2017) to attain semiparametric efficiency.
Chapter 1

Efficiency Bounds with Latent Variables

1.1 Introduction

I consider the problem of making inference about a \( p \)-dimensional parameter \( \theta \) which satisfies a set \( a \) of moment restrictions that are known to hold in the population. In contrast to conventional generalized method of moments (Hansen 1982), I allow these moment restrictions to contain random variables which are not observed by the researcher but require that the parameter \( \theta \) is point identified.

I consider both models defined by unconditional and conditional moment restrictions. The first model I consider is given by unconditional moment restrictions:

\[
E[g(U_t, \ldots, U_{t-l}, Z_t, \ldots, Z_{t-m}, \theta^0)] = 0.
\] (1.1)

where \( g \) is an \( H \)-dimensional vector of measurable functions that depends on the parameter \( \theta \in \Theta \subset \mathbb{R}^p \), of an unobserved random vector \( U_t \), of an observed random vector \( Z_t \), and (potentially) on lagged values of \( U_t \) and \( Z_t \). To economize notation in the sequel I will denote \( U_t = (U'_t, \ldots, U'_{t-l})' \), \( Z_t = (Z'_t, \ldots, Z'_{t-m})' \), and use \( (U, Z) \) to denote \( (U_t, Z_t) \) for a generic \( t \). Strict stationary of \( (U_t, Z_t) \) is assumed throughout.
The second model I consider is given by conditional moment restrictions:

\[ E[g(U_t, Z_t, \ldots, Z_{t-m}, \theta^0)|I_{t-1}] = 0. \] (1.2)

where the function \( g \) is as in model (1.1) and the information set \( I_{t-1} \) for conditioning is generated by the lagged values of \( Z_t \). It is worthwhile to remark that the relationship between models (1.1) and (1.2) is deeper than the obvious cosmetic similarities. Thanks to the law of iterated expectations, a set of conditional moment restrictions in the form of (1.2) may always be transformed into a set unconditional moment restrictions as for any conformable matrix of \( I_{t-1} \)-measurable random variables (instruments), say \( Z_{t-1} \):

\[ E[Z_{t-1}g(U_t, Z_t, \ldots, Z_{t-m}, \theta^0)] = 0. \] (1.3)

It will become apparent that the issue of efficiency estimation in model (1.2) amounts to characterizing the optimal set of instruments \( Z_{t-1} \) and suggesting a feasible procedure to estimate them. It should also be noted that model (1.2) includes, again by virtue of the law of iterated expectations, martingale difference sequences where the moment restriction implied by economic theory holds conditional on the lagged values of both \( U_t \) and \( Z_t \). In the sequel I will denote this information set as \( I^*_t \).

Since I set the focus here on efficient inference about \( \theta^0 \), I will consider the case where the moment restrictions (1.1) or (1.2) point identify the parameter of the model. The issue of when these models provide point identification is less obvious than in the hypothetical case where the latent variables were observed and the well understood classical theory of estimation with moment restrictions would apply. Indeed, it has been pointed out by Schennach (2014) that many partially identified models of interest may be written as particular cases of model (1.1). It will then be a maintained assumption throughout this work that standard conditions for the point identification of a parameter vector are satisfied for the fictitious circumstance where the latent variables are observed. Then the true parameter \( \theta^0 \) I hope to identify from the moment restrictions with latent variables is this same vector.

Many practical examples fall within the framework of models (1.1) and (1.2). I will consider as particular examples models of measurement error in affine regression, dynamic factor analysis
models, models of correlation between markets, and multivariate risk premium models.

Then the contribution of this chapter is to derive semiparametric efficiency bounds for models that contain latent variables. To derive bounds on asymptotic efficiency, I exhibit a parametric model consistent for each model (1.1) and (1.2) in which the inverse of the Fisher information can be seen as an efficiency bound. This approach to semiparametric efficiency extends the argument that Back and Brown (1992) used to to give a simple proof of the semiparametric efficiency of GMM with cross sectional data and no latent variables. Since I will then demonstrate an estimator that attains this efficiency bound in chapter 2, the parametric model I construct may be considered least favorable for the estimation of \( \theta \).

This chapter is organized as follows. Examples of econometric models that fall within the class of models I consider are discussed in section 1.2. The semiparametric efficiency bounds are derived in section 1.3. Section 1.4 concludes. All proofs appear in the appendix.

1.2 Examples

In this section I introduce specific examples which fit within the framework of models (1.1) and (1.2). The key feature of the models considered here is that they are defined by moment restrictions which contain latent variables and have parameters which are point identified. For many of the examples in this section I will suppress the intercept in linear models for notational convenience.

To appreciate the generality of the proposed framework, I consider a wide range of examples. The measurement error example in subsection 1.2.1 is used to illustrate the proposed method through Monte Carlo simulations and an empirical example in chapter 2.

1.2.1 Measurement error

As a first example, I consider a classical measurement error model where the regressor \( x_t^* \) is not observed, but a contaminated measurement (proxy) \( x_t \) is.

The model is summarized by the following equations

\[
\begin{align*}
y_t &= \beta x_t^* + u_t \\
x_t &= x_t^* + \epsilon_t
\end{align*}
\]
where the errors \((u_t, \epsilon_t)\) are uncorrelated with one another and there is no serial correlation in the error \(u_t\). It is well known that in the cross sectional setting the parameter \(\beta\) is not identified without some additional assumptions or data. A typical approach is to make use of a second proxy and estimate \(\beta\) using a linear IV estimator.

In the time series setting, provided that there is some serial correlation in the latent regressor \(x_t^*\), lagged values of contaminated measurement might be valid instruments. Provided that there is no serial correlation in \(\epsilon_t\) the researcher could estimate the parameter \(\beta\) by using \(x_{t-1}\) as an instrumental variable for \(x_t\). However, if the error \(\epsilon_t\) is serially correlated the exclusion restriction for instrumental variables estimation will be violated and the IV-estimator will not be consistent for the true parameter value.

It is perhaps surprising that in the case where there is serial correlation in \(\epsilon_t\) the parameter is still identified by observation of \((y_t, x_t)\) again provided that there is serial correlation in the latent regressor \(x_t^*\). The result, which appears to have originated with Gospodinov et al. (2017), shows that the parameter in this model is identified if there is no serial correlation in the error \(u_t\) but is serial correlation in the unobserved true regressor \(x_t^*\). In contrast to the IV estimator, the procedure that I will suggest consistently estimates \(\beta\) regardless of serial correlation in \(\epsilon_t\).

This model fits within the framework of model (1.1) with \(\theta = \beta\), \(Z_t = (y_t, y_{t-1}, x_t, x_{t-1})\), \(U_t = \ldots\).
constant returns to scale, perfect competition, and manager’s expectation of the marginal contribution of new capital goods to future profit, called q. In the q theory of investment the investment decisions of a firm are determined by the firm’s investment q. Investment q Theory

As an extension of the classical measurement error framework I consider the q theory of investment. In the q theory of investment the investment decisions of a firm are determined by the firm manager’s expectation of the marginal contribution of new capital goods to future profit, called marginal q. Hayashi (1982) has shown that constant returns to scale, perfect competition, and

Lemma 1.2.1. Suppose that \( E[x_t^* x_{t-1}^*] \neq 0 \). The parameter \( \theta = \beta \) is identified by the moment restrictions (1.4).

In chapter 2 I present Monte Carlo results for my estimation procedure using the moment restrictions (1.4). The Monte Carlo results suggest that the ELVIS estimator performs favorably relative to the estimator put forth by Gospodinov et al. (2017), particularly for values of \( \beta \) near zero. In addition, I estimate parameters from the long-run risk model of Bansal and Yaron (2004) using the classical measurement error framework.

Investment q Theory

As an extension of the classical measurement error framework I consider the q theory of investment. In the q theory of investment the investment decisions of a firm are determined by the firm manager’s expectation of the marginal contribution of new capital goods to future profit, called marginal q. Hayashi (1982) has shown that constant returns to scale, perfect competition, and

\[
g(U_t, Z_t, \theta) = \begin{pmatrix}
    u_t x_t^* \\
    u_t x_{t-1}^* \\
    u_{t-1} x_t^* \\
    u_t \epsilon_t \\
    u_{t-1} \epsilon_t \\
    u_t u_{t-1} \\
    x_t^* \\
    x_{t-1}^* \\
    \epsilon_t x_t^* \\
    \epsilon_{t-1} x_{t-1}^* \\
    \epsilon_t x_{t-1}^* \\
    \epsilon_{t-1} x_t^* \\
    \epsilon_t^2 - \epsilon_{t-1}^2 \\
    (x_t^*)^2 - (x_{t-1}^*)^2
\end{pmatrix} = \begin{pmatrix}
    (y_t - \beta x_t^*) x_t^* \\
    (y_t - \beta x_t^*) x_{t-1}^* \\
    (y_{t-1} - \beta x_{t-1}^*) x_t^* \\
    (y_{t-1} - \beta x_{t-1}^*) x_{t-1}^* \\
    (y_t - \beta x_t^*) (x_t - x_t^*) \\
    (y_t - \beta x_t^*) (x_{t-1} - x_{t-1}^*) \\
    (y_t - \beta x_t^*) (y_t - \beta x_{t-1}^*) \\
    x_t^* \\
    x_{t-1}^* \\
    (x_t - x_t^*) x_t^* \\
    (x_{t-1} - x_{t-1}^*) x_{t-1}^* \\
    (x_t - x_t^*) x_{t-1}^* \\
    (x_{t-1} - x_{t-1}^*) x_t^* \\
    (x_t^*)^2 - (x_{t-1}^*)^2 \\
    (x_t^*)^2 - (x_{t-1}^*)^2
\end{pmatrix}.
\]
efficient financial markets imply marginal q is equal to the ratio of the market valuation to the replacement value of capital, termed Tobin’s q. When these conditions are met the unobservable marginal q may be replaced by the observable Tobin’s q without consequence. However, if these economic assumptions fail and Tobin’s q is not a perfect proxy for marginal q the model may be analyzed in the context of the classical measurement error model. Assuming quadratic investment adjustment costs, the relationship between investment and marginal q takes the form of a linear regression function

\[ y_{ti} = x_{ti}^* \beta + u_{ti} \]

\[ x_{ti} = x_{ti}^* + \epsilon_{ti} \]

where \( y_{ti} \) is firm i’s investment in period \( t \), \( x_{ti}^* \) is firm i’s marginal q in period \( t \), and \( x_{ti} \) is Tobin’s q for firm i in period \( t \). Erickson and Whited (2000, 2011) have analyzed the econometrics of q theory as a classical measurement error problem. They propose to estimate the model using cross-section variation by adapting a result of Geary (1941). Their method is to strengthen the typical orthogonality restrictions that \( \text{Cov}(u_{ti}, x_{ti}^*) = 0 \) and \( \text{Cov}(\epsilon_{ti}, x_{ti}^*) = 0 \) to include higher order moments as

\[ \text{Cov}(u_{ti}, x_{ti}^{*2}) = 0 \]

\[ \text{Cov}(\epsilon_{ti}, x_{ti}^{*2}) = 0 \]

which imply, for \( 1 \leq j \leq n \)

\[ \beta = \frac{E(y_{tj}^2 x_{tj})}{E(y_{tj} x_{tj}^3)} \]

provided that \( \beta \neq 0 \) and \( E(x_{tj}^3) \neq 0 \).

Taking inspiration from the Erickson and Whited approach I may combine the higher order moments above with the restrictions of (1.4). The q theory model then fits within the framework of
model (1.1) with \( \theta = \beta \), \( Z_t = (y_t, y_{t-1}, x_t, x_{t-1}) \), \( U_t = (x_{t}^{\star}, x_{t-1}^{\star}) \) and moment restrictions given by

\[
g(U_t, Z_t, \theta) = \begin{pmatrix}
g_1(U_t, Z_t, \theta) \\
\vdots \\
g_n(U_t, Z_t, \theta)
\end{pmatrix}
\]

where:

\[
g_j(U_t, Z_t, \theta) = \begin{pmatrix}
u_{ij}x_{ij}^{\star} \\
u_{ij}x_{i-1j}^{\star} \\
u_{ij}x_j^{\star} \\
u_{ij}\epsilon_{ij} \\
u_{ij}\epsilon_{i-1j} \\
u_{ij}u_{ij} \\
\epsilon_{ij}x_{ij}^{\star} \\
\epsilon_{ij}x_{i-1j}^{\star} \\
\epsilon_{i-1j}x_j^{\star} \\
\epsilon_{i-1j}\epsilon_{ij} \\
\epsilon_{ij}^{2} - \epsilon_{i-1j}^{\star} \\
(x_{ij}^{\star})^2 - (x_{i-1j}^{\star})^2 \\
u_{ij}x_{ij}^{\star} \\
u_{ij}x_{ij}^{\star} \\
\epsilon_{ij}x_{ij}^{\star} \\
\epsilon_{ij}^{2} x_{ij}^{\star}
\end{pmatrix} = \begin{pmatrix}
(y_{ij} - \beta x_{ij}^{\star})x_{ij}^{\star} \\
(y_{ij} - \beta x_{ij}^{\star})x_{i-1j}^{\star} \\
(y_{ij} - \beta x_{i-1j}^{\star})x_{ij}^{\star} \\
(y_{ij} - \beta x_{ij}^{\star})(x_{ij} - x_{ij}^{\star}) \\
(y_{ij} - \beta x_{ij}^{\star})(x_{i-1j} - x_{ij}^{\star}) \\
(y_{ij} - \beta x_{i-1j}^{\star})(x_{ij} - x_{ij}^{\star}) \\
(y_{ij} - \beta x_{ij}^{\star})(y_{ij} - \beta x_{ij}^{\star}) \\
x_{ij}^{\star} \\
x_{ij}^{\star} \\
x_{i-1j}^{\star} \\
x_{i-1j}^{\star} \\
(x_{ij} - x_{ij}^{\star})^2 - (x_{ij} - x_{ij}^{\star})^2 \\
(x_{ij}^{\star})^2 - (x_{i-1j}^{\star})^2 \\
(y_{ij} - x_{ij}^{\star}\beta)x_{ij}^{\star} \\
(y_{ij} - x_{ij}^{\star}\beta)x_{ij}^{\star} \\
x_{ij} - x_{ij}^{\star} x_{ij}^{\star} \\
x_{ij} - x_{ij}^{\star} x_{ij}^{\star}
\end{pmatrix}
\]

### 1.2.2 Linear factor analysis

Several variations on models of linear factor analysis fit within my framework. To fix intuition, I first consider the simple linear factor analysis model
\[ y_t = B f_t + \epsilon_t \]  
\[ \text{(1.5)} \]

with

\[ \text{Var}(f_t) = I \]
\[ \text{Var}(\epsilon_t) = \Phi \]

where \( y_t \) is an \( n \times 1 \) vector of variables to be explained by a \( k \times 1 \) vector of latent factors \( f_t \), \( \epsilon_t \) is an \( n \times 1 \) mean zero error term which is uncorrelated with the latent factors \( f_t \), \( B \) is a \( n \times k \) vector of parameters, and \( \Phi \) is a diagonal positive definite matrix. This model has its origins in psychometrics (Spearman 1904) and is discussed extensively in, for example, Bartholomew (1987). The number of latent factors \( k \) is assumed to be smaller than the number of observables \( n \).

This model fits within the framework of model (1.1) with \( Z_t = y_t, U_t = f_t, \) and \( \theta = (\text{vec}(B)'', \text{diag}(\Phi)''')' \) where the moment restrictions are given by:

\[
g(U_t, Z_t, \theta) = \begin{pmatrix}
y_t \\
f_t \\
\text{vech} [\epsilon_t' \epsilon_t - \Phi] \\
\text{vech} [f_t f_t' - I] \\
\text{vec} [\epsilon_t f_t']
\end{pmatrix} = \begin{pmatrix}
y_t \\
f_t \\
\text{vech} [(y_t - B f_t)(y_t - B f_t)' - \Phi] \\
\text{vech} [f_t f_t' - I] \\
\text{vec} [(y_t - B f_t) f_t']
\end{pmatrix}. \tag{1.6}
\]

A reflection on equation (1.5) reveals that the factors \( f_t \) are not unique. Any invertible linear transformation of the factors \( f_t \) will satisfy (1.5) with a transformed parameter matrix. This lack of uniqueness means that the restriction \( \text{Var}(f_t) = I \) is simply a normalization. Even with the variance of factors normalized, the factors are still not defined without ambiguity as any orthogonal transformation of a set of factors will satisfy the normalization. Any identification results I present for factor analysis models accept this indeterminacy.

**Lemma 1.2.2.** Suppose that:

\[
\frac{n(n + 1)}{2} \geq n(k + 1).
\]
Up to orthogonal transformations of $f_t$ and the associated matrix of coefficients; the parameter vector $\theta = (\text{vec}(B)', \text{diag}(\Phi)')'$ is identified by the moment restrictions (1.6).

**Dynamic factor analysis model**

I now consider a dynamic factor analysis model where the latent variables exhibit serial correlation. Seminal references on dynamic factor analysis in econometrics include Geweke (1977) and Sargent et al. (1977). The model can be written using the following equations

\[
y_t = Bf_t + \epsilon_t \\
f_t = Af_{t-1} + u_t
\]

with

\[
\text{Var}(f_t) = I \\
\text{Var}(\epsilon_t) = \Phi
\]

where $\Phi$ is a diagonal positive definite matrix, $\epsilon_t$ is serially uncorrelated, $u_t, \epsilon_t$ are not correlated, and $u_t$ is uncorrelated from lagged values of $y_t$. The matrix of parameters $A$ is assumed to have eigenvalues within the unit circle. This model fits within the framework of model (1.1) with $Z_t = y_t$. 
\( \mathbf{U}_t = (f_t, f_{t-1}) \), and \( \theta = (\text{vec}(B)'', \text{vec}(A)'', \text{diag}(\Phi)''')' \) where the moment restrictions are given by:

\[
g(\mathbf{U}_t, \mathbf{Z}_t, \theta) = \begin{pmatrix}
    y_t \\
    f_t \\
    \text{vech}[(\epsilon_t' - \Phi)] \\
    \text{vech}[f_t f_t' - I] \\
    \text{vech}[f_{t-1} f_{t-1}' - I] \\
    \text{vec}[\epsilon_t f_t'] \\
    \text{vec}[u_t f_{t-1}'] \\
    \text{vec}[\epsilon_t' f_{t-1}'] \\
    \text{vec}[u_t y_{t-1}] \\
    \text{vec}[\epsilon_{t-1} f_{t-1}']
\end{pmatrix} = \begin{pmatrix}
    y_t \\
    f_t \\
    \text{vech}[(y_t - B f_t) (y_t - B f_t)' - \Phi] \\
    \text{vech}[f_t f_t' - I] \\
    \text{vech}[f_{t-1} f_{t-1}' - I] \\
    \text{vec}[(y_t - B f_t) f_t'] \\
    \text{vec}[(y_t - B f_t) f_{t-1}'] \\
    \text{vec}[(y_t - B f_t) (y_{t-1} - B f_{t-1})'] \\
    \text{vec}[(y_t - B f_t) y_{t-1}'] \\
    \text{vec}[(y_{t-1} - B f_{t-1}) f_{t-1}']
\end{pmatrix}. \tag{1.7}
\]

**Lemma 1.2.3.** Suppose that the matrices \( A \) and \( B \) have full column rank and:

\[
\frac{n(n+1)}{2} \geq n(k+1).
\]

Up to orthogonal transformations of \( f_t \) and the associated matrices of coefficients; the parameter vector \( \theta = (\text{vec}(A)'', \text{vec}(B)'', \text{diag}(\Phi)''')' \) is identified by the moment restrictions above.

**Contagion**

The factor structure of the previous section may be augmented to model correlation between markets. Suppose \( y_t \) is a return vector of market indexes for different stock exchanges. Comprehensive treatments of contagion modeling in finance are available in Dungey et al. (2010) and Darolles and Gouriéroux (2015).

For the exposition here I limit my examination to the model put forth by Gagliardini and Gouriéroux (2016). The dependence between markets can be analyzed using a dynamic model of the form:

\[
y_t = B f_t + C y_{t-1} + \epsilon_t.
\]
The matrix of parameters $C$ is interpreted as a measure of the interdependence between markets. This model fits within the framework of model (1.1) with $Z_t = (y_t', y_{t-1}')'$, $U_t = f_t$, and $\theta = (\text{vec}(B)', \text{vec}(C)', \text{diag}(\Phi)')'$ where the moment restrictions are given by:

$$
g(U_t, Z_t, \theta) = \begin{pmatrix}
  y_t \\
  f_t \\
  \text{vech}[\epsilon_t' \epsilon_t' - \Phi] \\
  \text{vech}[f_t f_t' - I] \\
  \text{vec}[\epsilon_t f_t'] \\
  \text{vec}[\epsilon_t y_{t-1}'] \\
  \text{vec}[f_t y_{t-1}']
\end{pmatrix}, \quad (1.8)
$$

$$
g(U_t, Z_t, \theta) = \begin{pmatrix}
  y_t \\
  f_t \\
  \text{vech}[(y_t - B f_t - C y_{t-1})(y_t - B f_t - C y_{t-1})' - \Phi] \\
  \text{vech}[f_t f_t' - I] \\
  \text{vec}[(y_t - B f_t - C y_{t-1}) f_t'] \\
  \text{vec}[(y_t - B f_t - C y_{t-1}) y_{t-1}'] \\
  \text{vec}[f_t y_{t-1}']
\end{pmatrix}. \quad (1.9)
$$

The following lemma shows that the parameters of this model are identified by the moment restrictions (1.9).

**Lemma 1.2.4.** Suppose that:

$$
\frac{n(n+1)}{2} \geq n(k+1).
$$

Up to orthogonal transformations of $f_t$ and the associated matrix of coefficients; the parameter vector $\theta = (\text{vec}(C)', \text{vec}(B)', \text{diag}(\Phi)')'$ is identified by the moment restrictions (1.9).

The contagion model may be augmented to include dynamic factors as for example:

$$
y_t = B f_t + C y_{t-1} + \epsilon_t \\
f_t = A f_{t-1} + u_t.
$$
Darolles et al. (2014) have shown that the contagion matrix is identified in this case under additional restrictions which will not be considered in detail here.

1.2.3 Multivariate volatility model

The multivariate volatility model with risk premium presented in Gagliardini and Gouriéroux (2016) can be written as

\[ y_t = a + B \text{vech}(\Sigma_t) + \Sigma_t^{1/2} \epsilon_t \]

where \( \epsilon_t \) is independent and identically distribution random variable with variance one, \( y_t \) is a vector of asset returns, and \( \Sigma_t \) is the unobserved stochastic volatility-covolatility matrix. This model fits within the framework of model (1.2) with \( \theta = (a', \text{vec}(B')', Z_t = y_t, U_t = \text{vech}(\Sigma_t), \) and

\[
g(U_t, Z_t, \theta) = \begin{pmatrix} y_t - a - B \text{vech}(\Sigma_t) \\ \text{vech}(y_t y_t') - \text{vech}(\Sigma_t) \end{pmatrix}. \]

1.3 Efficiency in moment restriction models

In this section I will derive a semiparametric efficiency bounds for the models (1.1) and (1.2). To derive the efficiency bounds I construct parametric models that are nested within models (1.1) and (1.2) and compute their Fisher information. Since each parametric model is nested within a semiparametric model, it is clear that no semiparametric procedure could have asymptotic variance smaller than the inverse of the corresponding Fisher information.

Throughout this section I will refer to any matrix \( A \) an efficiency bound if for any asymptotically normal regular estimator of \( \theta_0 \)

\[ \text{Var}(\hat{\theta}) - A \]

is positive definite. I will call the matrix \( A \) a feasible efficiency bound if it is an efficiency bound and there exists an asymptotically normal regular estimator of \( \theta_0 \) with:

\[ \text{Var}(\hat{\theta}) = A. \]
1.3.1 Intuition

In this subsection I will develop the intuition for the efficiency bounds that apply to models (1.1) and (1.2). First I will consider model (1.1). There are $H$ moment restrictions

$$E[g(U, Z, \theta)] = 0$$

where $\theta \in \Theta \subset \mathbb{R}^p$. The set $\Theta$ is such that

$$\{E[g(U, Z, \theta)] = 0, \theta \in \Theta\} \iff \theta = \theta^0$$

where the true unknown value $\theta^0$ belongs to the interior of $\Theta$. I assume that the function $g(u, z, \cdot)$ is continuously differentiable on the interior of $\Theta$ for any possible values $(u, z)$ of the random vector $(U, Z)$. This differentiability assumption excludes moment functions with cusps and discontinuities such as those that arise in quantile regression models.

As in the familiar case of estimation using dependent data without latent variables, the long run variance of the moment restrictions will play a role in the efficiency bound I will derive. I will use the following expression for the long run variance of the moment restrictions at the true value of the parameter and maintain the assumption that this matrix is non-singular:

$$\Sigma_{LR} = \text{Cov} \left( g(U_t, Z_t, \theta^0), \sum_{s=-\infty}^{\infty} g(U_s, Z_s, \theta^0) \right).$$

Since estimation will be based on the observation of a serially dependent sequence $z_1, z_2, \ldots, z_T$ while $U$ remains latent, the following decomposition of the long run variance matrix for the moment restrictions will be of use:

$$\Sigma_{LR} = \Sigma^L_E + \Sigma^L_V$$

$$\Sigma^L_E = E \left[ \text{Cov} \left( g(U_t, Z_t, \theta^0), \sum_{s=-\infty}^{\infty} g(U_s, Z_s, \theta^0) | Z_t \right) \right]$$

$$\Sigma^L_V = \text{Cov} \left( E(g(U_t, Z_t, \theta^0) | Z_t), \sum_{s=-\infty}^{\infty} E(g(U_s, Z_s, \theta^0) | Z_t) \right).$$

I note here that both $\Sigma^L_V$ and $\Sigma^L_E$ may be of reduced rank even when $\Sigma_{LR}$ is nonsingular.
The simplest intuition leading to an efficiency bound is that the law of iterated expectations allows me to rewrite the moment restrictions as:

\[ E[g(U, Z, \theta)] = E[\tilde{g}^0(Z, \theta)] \]
\[ \tilde{g}^0(Z, \theta) = E[g(U, Z, \theta)|Z]. \]

Thus the moment restrictions now read

\[ E[\tilde{g}^0(Z, \theta)] = 0 \]

and the optimal variance for GMM estimation from Hansen (1982) is

\[ [\Gamma^0_V (\Sigma^L VR^0_V)^{-1} \Gamma^0_V ]^{-1} \] (1.10)

where:

\[ \Gamma^0_V = E \left[ \frac{\partial \tilde{g}^0(Z, \theta^0)}{\partial \theta'} \right]. \] (1.12)

The efficiency bound considered in (1.10) is actually a generalization of the classical efficiency bound for GMM estimation that accommodates the possibility that the variance of the moment restrictions, \( \Sigma^L VR \) is of reduced rank. Inspired by a result from Peñaranda and Sentana (2012), replacing the inverse of \( \Sigma^L VR \), which may not exist, with its Moore-Penrose generalized inverse \((\Sigma^L VR)^+\) results in an appropriate expression for the efficiency bound.

However, I do not expect that the efficiency bound in (1.10) is feasible. As the variable \( U \) is unobserved, the researcher cannot hope to identify the conditional distribution of \( U \) given \( Z \) and therefore cannot compute the functions \( \tilde{g}^0(Z, \theta^0) \). I will show that a feasible efficiency bound can be written

\[ EFF = [\Gamma^0_V (\Sigma^L VR^0_V)^{-1} \Gamma^0_V ]^{-1} \] (1.11)
\[ \Gamma_V = E \left[ \frac{\partial \tilde{g}_\omega(Z, \theta^0)}{\partial \theta'} \right] \] (1.12)

where \( \tilde{g}_\omega(z, \theta) \) is a function which will be defined in the sequel that coincides with the true function
\tilde{g}^0(z, \theta) \text{ for } \omega = \theta = \theta^0;

\tilde{g}_{\theta^0}(Z, \theta^0) = a_s \tilde{g}^0(Z, \theta^0).

but may exhibit different behavior in the neighborhood of \( \theta^0 \). The crucial intuition is that I will consider a function

\[ \tilde{g}_\omega(Z, \theta) = \tilde{E}_\omega[g(U, Z, \theta) | Z], \omega = \theta \]

where \( \tilde{E}_\omega[|Z] \) denotes a conditional expectation operator given \( Z \) that is computed using a tilted conditional distribution of \( U \) given \( Z \). This tilted distribution depends on \( \theta \); I have used the notation \( \tilde{g}_\omega(Z, \theta) \) with \( \omega = \theta \) to emphasize that \( \theta \) enters the computation of this function in two distinct ways. Since the true conditional distribution is not identified, I will characterize an efficiency bound by constructing a family of conditional distributions of \( U \) given \( Z \) that does not artificially simplify the estimation of \( \theta^0 \). That is, the family of conditional distributions must not provide any information which can be used to identify the parameter vector beyond what is contained in the moment restrictions. In the sequel I will call a family of conditional distributions with this property least favorable for the estimation of \( \theta^0 \).

I now consider the derivation of an efficiency bound for model (1.2). In this case, I have \( H \) conditional moment restrictions

\[ E[g(U_t, Z_t, \theta^0) | I_{t-1}] = 0 \quad (1.13) \]

where the conditioning information \( I_{t-1} \) is the sigma-algebra generated by the lagged values of \( Z_t \). Again, the parameter is \( \theta \in \Theta \subset \mathbb{R}^p \) and the set \( \Theta \) is such that

\( \{ E[g(U_t, Z_t, \theta^0) | I_{t-1}] = 0 | \theta \in \Theta \} \leftrightarrow \theta = \theta^0 \)

where the unknown value \( \theta^0 \) belongs to the interior of \( \Theta \). The function \( g(u, z, \cdot) \) is continuously differentiable on the interior of \( \Theta \) for any possible values \( (u, z) \) of the random vector \( (U, Z) \). In addition, the conditional variance of the moment restrictions at the true parameter value \( \theta^0 \) is
non-singular with probability one:

\[ \Sigma^{t-1} = \text{Var}[g(U_t, Z_t, \theta^0) | I_{t-1}] . \]

Since estimation will be based on the observation of a serially dependent sequence \( z_1, z_2, \ldots, z_T \) while \( U \) remains latent, I will make use of the following decomposition of the conditional variance matrix for the moment restrictions

\[
\Sigma^{t-1} = \Sigma^t_E + \Sigma^t_V = \Sigma_{t-1}^E + \Sigma_{t-1}^V \\
\Sigma^t_E = E \left[ \text{Var} \left( g(Z_t, U_t, \theta^0) | I_{t-1}, Z_t \right) | I_{t-1} \right] \\
\Sigma^t_V = \text{Var} \left( E \left( g(Z_t, U_t, \theta^0) | I_{t-1}, Z_t \right) | I_{t-1} \right) \\
\]

where \( \Sigma^t_E \) and \( \Sigma^t_V \) may be of reduced rank.

Thanks to the law of iterated expectations, the main intuition is essentially unchanged from the unconditional case. I consider eliminating the latent variable using the law of iterated expectations as:

\[
E[g(U_t, Z_t, \theta) | I_{t-1}] = E[g_0^{t-1}(Z_t, \theta) | I_{t-1}] \\
g_0^{t-1}(Z_t, \theta) = E[g(U_t, Z_t, \theta) | I_{t-1}, Z_t] \\
= E[g(U_t, Z_t, \theta) | I_t].
\]

So the moment restricts can be written

\[
E[g_0^{t-1}(Z_t, \theta) | I_{t-1}] = 0
\]

and the classic semiparametric efficiency bound of Chamberlain (1987) is:

\[
\{E[\Gamma_0^{t-1}(\Sigma^t_V^{-1} + \Sigma^t_V)^{-1}] \}^{-1}
\]
where:

\[ \Gamma_0^{t-1} = E \left[ \frac{\partial \tilde{g}_t^{0}(Z, \theta^0)}{\partial \theta} \big| I_{t-1} \right]. \] (1.15)

up to the inclusion of a Moore-Penrose inverse to accommodate the possible singularity of \( \Sigma_V^{t-1} \).

However, the efficiency bound in (1.14) is probably too optimistic. Again, without observing the variable \( U_t \), the conditional distribution of \( U_t \) given \( I_t \) cannot be identified and the moment functions \( \tilde{g}_t^{0}(Z, \theta) \) cannot be computed. Instead a feasible efficiency bound is given by

\[ \{ E[\Gamma_{t-1}^{0}(\Sigma_V^{t-1})^\dagger \Gamma_{t-1}^{0}] \}^{-1} \] (1.16)

\[ \Gamma_{t-1}^{0} = E \left[ \frac{\partial \tilde{g}_{t-1}^{0}(Z, \theta^0)}{\partial \theta} \big| I_{t-1} \right] \] (1.17)

where \( \tilde{g}_{t-1, \omega}(Z, \theta) \) is a function that will be explicitly defined later in this section. This function coincides with the true function \( \tilde{g}_{t-1}(Z, \theta) \) when \( \theta = \omega = \theta^0 \)

\[ \tilde{g}_{t-1}^{0}(Z, \theta^0) = \text{a.s.} \tilde{g}_{t-1, \theta^0}(Z, \theta^0) \]

but may exhibit different behavior in the neighborhood of \( \theta^0 \). The main intuition is that I consider a function

\[ \tilde{g}_{t-1, \omega}(Z, \theta) = \tilde{E}_\omega[g(U_t, Z_t, \theta)|Z_t, I_{t-1}], \omega = \theta \]

where \( \tilde{E}_\omega[|Z_t, I_{t-1}] \) is a conditional expectation operator given \((Z_t, I_{t-1})\) that is again computed using a family of tilted conditional distributions of \( U_t \) given \((Z_t, I_{t-1})\) which is least favorable for the estimation of \( \theta^0 \).

1.3.2 Preliminaries

In this section I will introduce the assumptions used to derive my efficiency bounds and transform the moment function into a form that will simplify the exposition in the sequel. For model (1.1), the moment function is given by \( g(U_t, Z_t, \theta^0) \). I denote the contemporaneous conditional variance
\[ \Sigma_E = E \left[ Var \left( g(U_t, Z_t, \theta^0) | Z_t \right) \right]. \]

Let the rank of the matrix \( \Sigma_E \) be denoted by \( r_E \).

For model (1.2), the moment function is given by \( g(U_t, Z_t, \theta^0) \). I denote the rank of the matrix \( \Sigma_{E}^{t-1} \) by \( r_E^{t-1} \) and note that:

\[ r_E^{t-1} = rank(\Sigma_E^{t-1}) \leq H = rank(\Sigma^{t-1}). \]

The superscript \( t-1 \) is included here to emphasize that the rank of \( \Sigma_{E}^{t-1} \) may be dependent on the lagged values of \( Z_t \).

The following Propositions rearrange the moment functions by means of linear transformations so that the matrices \( \Sigma_E \) or \( \Sigma_{E}^{t-1} \) are block diagonal.

**Proposition 1.3.1.** There exists an orthogonal matrix \( B \) of dimension \( H \), such that

\[
Bg(U_t, Z_t, \theta) = \begin{pmatrix}
B_1g(U_t, Z_t, \theta) \\
B_2g(U_t, Z_t, \theta)
\end{pmatrix}
= \begin{pmatrix}
g_1(U_t, Z_t, \theta) \\
h_2(U_t, Z_t, \theta)
\end{pmatrix}
= g^*(Z_t, U_t, \theta)
\]

where for a diagonal positive definite matrix \( \Lambda \) of rank \( r_E \):

\[
E \left[ Var \left( g_1^*(U_t, Z_t, \theta^0) | Z_t \right) \right] = \Lambda
\]
\[
E \left[ Var \left( h_2(U_t, Z_t, \theta^0) | Z_t \right) \right] = 0.
\]

**Proposition 1.3.2.** There exists an \( I_{t-1} \)-measurable orthogonal matrix \( B^{t-1} \) of dimension \( H \), such
that

\[
B^{t-1}g(U_t, Z_t, \theta) = \begin{pmatrix}
B_1^{t-1}g(U_t, Z_t, \theta) \\
B_2^{t-1}g(U_t, Z_t, \theta)
\end{pmatrix}
= \begin{pmatrix}
g_{1,t-1}^*(U_t, Z_t, \theta) \\
h_{2,t-1}(U_t, Z_t, \theta)
\end{pmatrix}
= g_{t-1}^*(U_t, Z_t, \theta)
\]

where for a diagonal positive definite matrix \(\Lambda^{t-1}\) of rank \(r^{t-1}\):

\[
E\left[Var \left( g_{1,t-1}^*(U_t, Z_t, \theta^0) | Z_t, I_{t-1} \right) | I_{t-1} \right] = \Lambda^{t-1}
\]

\[
E\left[Var \left( h_{2,t-1}(U_t, Z_t, \theta^0) | Z_t, I_{t-1} \right) | I_{t-1} \right] = 0.
\]

The lemmata below deduce from Propositions 1.3.1 and 1.3.2 that \(h_2(U_t, Z_t, \theta^0)\) and \(h_{2,t-1}(U_t, Z_t, \theta^0)\) do not depend on the latent variables.

**Lemma 1.3.1.** \(h_2(U_t, Z_t, \theta^0)\) is \(Z_t\)-measurable.

**Lemma 1.3.2.** \(h_{2,t-1}(U_t, Z_t, \theta^0)\) is \(I_t\)-measurable.

Assumptions 1.3.1 and 1.3.2 are used to derive the efficiency bound for model (1.1).

**Assumption 1.3.1.** \(E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta^0) \right] \) is full-column rank

**Assumption 1.3.2.** The long run variance \(\Sigma^{LR}\) is non-singular.

Assumption 1.3.2 implies that for sufficiently large \(m\) the matrix

\[
\Sigma_{22}^m = Cov \left( h_2(U_t, Z_t, \theta^0), \sum_{s=t-m}^{t+m} h_2(U_s, Z_s, \theta^0) \right)
\]

is non-singular.

Assumptions 1.3.3 and 1.3.4 are used to derive the efficiency bound for model (1.2).

**Assumption 1.3.3.** The conditional variance matrix \(\Sigma^{t-1}\) is non-singular with probability one.
Assumption 1.3.4. The matrix:

\[ E \left\{ E \left[ \frac{\partial}{\partial \theta'} h_{2,t-1}(U_t, Z_t, \theta^0) \right| I_{t-1} \right] Var(h_{2,t-1}(U_t, Z_t, \theta^0) \mid I_{t-1})^{-1} E \left[ \frac{\partial}{\partial \theta'} h_{2,t-1}(U_t, Z_t, \theta^0) \right| I_{t-1} \right] \}
\]

is non-singular.

Assumptions 1.3.1-1.3.4 are local identification conditions required to derive asymptotic efficiency bounds and Gaussian limiting distributions. The conditions for local identification take the form of full rank and nonsingularity requirements. These conditions appear familiar to those seen in all forms of extremum estimation. However, the conditions specifically apply to “important directions” of the vector valued moment function rather than the function as a whole. The important direction of the moment functions are given by \( h_2 \) and \( h_{2,t-1} \) for unconditional and conditional models respectively. The interpretation of these directions is given by lemmata 1.3.1 and 1.3.2; the relevant directions for local identification are those which do not depend on the latent variables when \( \theta = \theta^0 \).

Using Propositions 1.3.1 and 1.3.2 I obtain orthogonal matrices \( B, B_{t-1} \) such that

\[ B \Sigma_E B' = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \]

\[ B_{t-1} \Sigma_{t-1} B_{t-1}' = \begin{pmatrix} \Lambda_{t-1} & 0 \\ 0 & 0 \end{pmatrix} \]

Furthermore, I define the matrices, \( \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \Sigma_{11}^{t-1}, \Sigma_{12}^{t-1} \) and \( \Sigma_{22}^{t-1} \) through the partitioned matrix equations

\[ B \Sigma_{L \bar{V}} B' = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \]

\[ B_{t-1} \Sigma_{t-1} B_{t-1}' = \begin{pmatrix} \Sigma_{t-1}^{11} & \Sigma_{t-1}^{12} \\ \Sigma_{t-1}^{12}' & \Sigma_{t-1}^{22} \end{pmatrix} \]

where \( \Sigma_{11} \) is \( r_E \times r_E \) and \( \Sigma_{11}^{t-1} \) is \( r_E^{t-1} \times r_E^{t-1} \).

By virtue of the lemmata 1.3.1 and 1.3.2, Assumption 1.3.2 implies that \( \Sigma_{22} \) is nonsingular, and Assumption 1.3.3 implies that \( \Sigma_{22}^{t-1} \) is nonsingular with probability one. I define another set of
invertible linear transformations $g^* \rightarrow h$ and $g_{t-1}^* \rightarrow h_{t-1}$ by the matrices $M, M_{t-1}^{-1}$:

$$M = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix},$$

$$M_{t-1}^{-1} = \begin{pmatrix} I & -\Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1} \\ 0 & I \end{pmatrix}.$$  

$$Mg^*(U_t, Z_t, \theta) = h(U_t, Z_t, \theta)$$

$$= \begin{pmatrix} h_1(U_t, Z_t, \theta) \\ h_2(U_t, Z_t, \theta) \end{pmatrix}$$

$$M_{t-1}^{-1}g^*_{t-1}(U_t, Z_t, \theta) = h(U_t, Z_t, \theta)$$

$$= \begin{pmatrix} h_{1,t-1}(U_t, Z_t, \theta) \\ h_{2,t-1}(U_t, Z_t, \theta) \end{pmatrix}.$$  

By the linearity of the expectation operator

$$E(g(U_t, Z_t, \theta)) = 0 \iff E(h(U_t, Z_t, \theta)) = 0$$

$$E(g(U_t, Z_t, \theta)|I_{t-1}) = 0 \iff E(h_{t-1}(U_t, Z_t, \theta)|I_{t-1}) = 0$$

so that the transformed moment functions given by $h, h_{t-1}$ are equivalent to those given in the models (1.1) and (1.2). Collecting these results I have the follow facts about the transformed
moment function $h, h_{t-1}$

$$
\Sigma_E^h = E \left[ Var \left( h(U_t, Z_t, \theta^0) \right) \mid Z_t \right]
= \begin{pmatrix}
\Lambda & 0 \\
0 & 0 
\end{pmatrix}
$$

$$
\Sigma_{V}^{L,R,h} = Cov \left( E(h(U_t, Z_t, \theta^0) \mid Z_t), \sum_{s=-\infty}^{\infty} E(h(U_s, Z_s, \theta^0) \mid Z_t) \right)
= \begin{pmatrix}
F_1 & 0 \\
0 & \Sigma_{22}
\end{pmatrix}
$$

$$
\Sigma_{E}^{h,t-1} = E \left[ Var \left( h_{t-1}(U_t, Z_t, \theta^0) \mid Z_t, I_{t-1} \right) \mid I_{t-1} \right]
= \begin{pmatrix}
\Lambda^{t-1} & 0 \\
0 & 0 
\end{pmatrix}
$$

$$
\Sigma_{V}^{h,t-1} = Var[E(h_{t-1}(U_t, Z_t, \theta^0) \mid Z_t, I_{t-1}) \mid I_{t-1}]
= \begin{pmatrix}
F_1^{t-1} & 0 \\
0 & \Sigma_{22}^{t-1}
\end{pmatrix}
$$

where $F_1$ is a positive semidefinite matrix, and $F_1^{t-1}$ is a positive semidefinite matrix with probability one. In the sequel I find it convenient to work with the transformed moment restrictions

$$
E(h(U_t, Z_t, \theta)) = 0
$$

$$
E(h_{t-1}(U_t, Z_t, \theta) \mid I_{t-1}) = 0
$$

rather than the equivalent conditions in (1.1) or (1.2). This is merely for theoretical exposition and does not affect the estimation procedures I suggest in any way. Indeed the transformations are infeasible as they are dependent on decompositions of variance matrices which cannot be estimated from data that does not include the latent variables.

1.3.3 Nested parametric models

I take inspiration from Back and Brown (1992) and build parametric models around the true data generating process (DGP) for which the Fisher information can be computed. I can therefore
deduce the asymptotic variance of the maximum likelihood estimator for the parametric problem given by the models I construct. Since these parametric models are nested in the moment restriction models (1.1) and (1.2), I obtain valid semiparametric efficiency bounds for models (1.1) and (1.2).

The DGP for any sample is characterized by the true unknown joint distribution of \((U_s, Z_s)_{s=1}^T = (U_1^T, Z_1^T)\) which I denote by \(P_{1,0(U,Z)}^{1,T}\). The parametric models that I will consider are defined as Radon-Nikodym derivatives with respect to probability measures induced from this DGP.

**Parametric model for the unconditional model**

The joint distribution is decomposed into two factors: A marginal probability distribution for \(Z_1^T\) denoted by \(P_{1,0,Z}^{1,T}\) and a family of conditional probability distributions for \(U_1^T\) given \(Z_1^T = z_1^T\) denoted by \(P_{1,0,U|Z}^{1,T}\).

I will define a parametric model \(P_{\theta_0, (U,Z)}^{1,T}\), \(\theta \in N(\theta_0) \subset \hat{\Theta}\), around the DGP, meaning that I require:

\[
P_{\theta_0, (U,Z)}^{1,T} = P_{0, (U,Z)}^{1,T}.
\]

The parametric model I will consider is defined by the following probability density functions

\[
\frac{dP_{\theta_0, (U,Z)}^{1,T}(u_1^T | Z_1^T = z_1^T)}{dP_{0, (U,Z)}^{1,T}(u_1^T | Z_1^T = z_1^T)} \propto \exp \left( \sum_{j=1}^T \lambda(\theta)' h_1(u_j, z_j, \theta^0) \right) P_0(du_1^T | Z_1^T = z_1^T)
\]

and

\[
\frac{dP_{\theta_0, (Z)}^{1,T}(z_1^T)}{dP_{0,Z}^{1,T}(z_1^T)} \propto \exp \left( \nu(\theta)' \sum_{j=1-m}^{T+m} h_2(u_j, z_j, \theta^0) \right) P_0(dz_{1-m}^T, dz_{T+1}^{T+m}, du_{1-m}^{T+m} | Z_1^T = z_1^T)
\]

where \(P_0\) is a probability measure given by the DGP on random variables which are apparent from the context. Moreover, the functions \(\lambda(\cdot)\) and \(\nu(\cdot)\) are defined on a neighborhood \(\theta \in N(\theta^0) \subset \hat{\Theta}\) of \(\theta^0\), take their values in \(\mathbb{R}^{r_E}\) and \(\mathbb{R}^{H-r_E}\) respectively, and are such that:

\[
\lambda(\theta^0) = 0
\]

\[
\nu(\theta^0) = 0.
\]
This model is indexed by the nonnegative integer \( m \).

Expectations with respect to the parametric model can now be defined as

\[
\tilde{E}_\theta(q(U, Z)) = \int q(u, z) P_{\theta, U | Z}(du | z) P_{\theta, Z}(dz) \tag{1.18}
\]

and

\[
\tilde{E}_\theta(q(U, Z) | Z = z) = \int q(u, z) P_{\theta, U | Z}(du | z).
\]

To economize notations in the sequel I will denote the tilted moment functions as follows:

\[
\tilde{g}_\theta(z, \theta) = \tilde{E}_\theta(g(U, Z, \theta) | Z = z)
\]

\[
\tilde{g}_\theta^*(z, \theta) = \tilde{E}_\theta(g^*(U, Z, \theta) | Z = z)
\]

\[
\tilde{h}_\theta(z, \theta) = \tilde{E}_\theta(h(U, Z, \theta) | Z = z).
\]

Recall that the function \( \tilde{g}_\theta(z, \theta) \) has already been introduced in Section 3.1. The existence of the functions \( \lambda(\cdot) \) and \( \nu(\cdot) \) is assured by the following proposition.

**Proposition 1.3.3.** Suppose that \( m \) is sufficiently large that \( \Sigma_{m}^{-1} \) is invertible. There exists a neighborhood \( N(\theta^0) \) of \( \theta^0 \) and functions \( \lambda(\cdot) \) and \( \nu(\cdot) \) from \( N(\theta^0) \) to \( \mathbb{R}^{r_E} \) and \( \mathbb{R}^{H-r_E} \) respectively such that \( \lambda(\theta^0) = 0, \nu(\theta^0) = 0 \), and for all \( \theta \in N(\theta^0) \)

\[\tilde{E}_\theta(g(U, Z, \theta)) = 0\]

where \( \tilde{E}_\theta(\cdot) \) is defined in (1.18). Furthermore,

\[
\frac{\partial}{\partial \theta} \nu'(\theta^0) = -E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]' (\Sigma_{m}^{-1})^{-1}.
\]

Proposition 1.3.3 shows that the parametric model introduced in this section is nested within model (1.1).
Parametric model for the conditional model

I will now develop an analog for Proposition 1.3.3 for the conditional moment restriction model. The result is again a parametric model with an exponential form that is defined through an application of the implicit function theorem. The parametric model I wish to consider is given by:

\[
\frac{dP_\theta(u_0, z_0)}{dP_0(u_0, z_0)} = 1
\]

\[
\frac{dP_\theta(u_t | z_t, I^*_t)}{dP_0(u_t | z_t, I^*_t)} = \frac{\exp(\lambda_{t-1}(\theta)' h_{1,t-1}(u_t, z_t, \theta_0))}{\int \exp(\lambda_{t-1}(\theta)' h_{1,t-1}(u_t, z_t, \theta_0)) P_0(du_t | z_t, I^*_t)}
\]

\[
\frac{dP_\theta(z_t | I^*_t)}{dP_0(z_t | I^*_t)} = \frac{\exp(\nu_{t-1}(\theta)' h_{2,t-1}(u_t, z_t, \theta_0))}{\int \exp(\nu_{t-1}(\theta)' h_{2,t-1}(u_t, z_t, \theta_0)) P_0(dz_t | I^*_t)}
\]

where \( P_0 \) is a probability measure given by the DGP on random variables which are apparent from the context and \( I^*_t \) is the sigma-algebra generated by the lagged values of \((Z_t, U_t)\). The functions \( \lambda_{t-1}(\cdot) \) and \( \nu_{t-1}(\cdot) \) are defined on a neighborhood \( \theta \in N(\theta_0) \subset \Theta \) of \( \theta_0 \), take their values in \( \mathbb{R}^{r_{t-1}} \) and \( \mathbb{R}^{H-r_{t-1}} \) respectively, and are such that:

\[
\lambda_{t-1}(\theta_0) = 0
\]

\[
\nu_{t-1}(\theta_0) = 0.
\]

Conditional expectations with respect to the parametric model can now be defined as

\[
\tilde{E}_\theta(q(U_t, Z_t) | I_{t-1}) = \int q(u_t, z_t) P_\theta(du_t | z_t, I_{t-1}) P_\theta(dz_t | I_{t-1})
\]

(1.19)

and

\[
\tilde{E}_\theta(q(U_t, Z_t) | Z_t = z_t, I_{t-1}) = \int q(u_t, z_t) P_\theta(du_t | z_t, I_{t-1}).
\]
To economize notations in the sequel I will denote the tilted moment functions as follows:

\[
\tilde{g}_{\theta,t-1}(z_t, \theta) = \tilde{E}_{\theta}(g(U_t, Z_t, \theta) | Z_t = z_t, I_{t-1})
\]

\[
\tilde{g}^*_{\theta,t-1}(z_t, \theta) = \tilde{E}_{\theta}(g^*(U_t, Z_t, \theta) | Z_t = z_t, I_{t-1})
\]

\[
\tilde{h}_{\theta,t-1}(z_t, \theta) = \tilde{E}_{\theta}(h(U_t, Z_t, \theta) | Z_t = z_t, I_{t-1}).
\]

**Proposition 1.3.4.** There exists a neighborhood \(N(\theta^0)\) of \(\theta^0\) and functions \(\lambda_{t-1}()\) and \(\nu_{t-1}()\) from \(N(\theta^0)\) to \(\mathbb{R}^{r_t-1}\) and \(\mathbb{R}^{H-r_t-1}\) respectively such that \(\lambda_{t-1}(\theta^0) = 0\), \(\nu_{t-1}(\theta^0) = 0\), and for all \(\theta \in N(\theta^0)\)

\[
\tilde{E}_{\theta}(g(U_t, Z_t, \theta) | I_{t-1}) = 0
\]

where \(\tilde{E}_{\theta}()\) defined in (1.19). Furthermore,

\[
\frac{\partial}{\partial \theta'} \nu_{t-1}(\theta^0) = -E \left[ \frac{\partial}{\partial \theta'} h_{2,t-1}(U_t, Z_t, \theta) | \theta = \theta^0, I_{t-1} \right] \left( \Sigma_{22}^{-1} \right)^{-1}.
\]

Proposition 1.3.4 shows that the parametric model introduced in this section is nested within model (1.2).

### 1.3.4 Fisher information and efficiency

I now derive efficiency bounds using the parametric models of the previous subsection. The efficiency bounds are determined by computing the asymptotic variance of the maximum likelihood estimators for the parametric problem given by the model defined in Propositions 1.3.3 and 1.3.4. I will propose feasible estimators to attain these bounds in the coming sections.

This argument is a generalization of Back and Brown (1992) that accommodates dependent data and latent variables. In Proposition 1.3.3 I have defined a joint density for \((U^T, Z^T)\) with respect to the true DPG. To get the likelihood function for the observables I will have to integrate out the latent variables. I therefore set the focus on the marginal distribution of \((Z_t)^T_{t=1}\):

\[
\frac{dP_{0,Z}^T(z^T_1)}{dP_{0,Z}^T(z^T_1)} = \frac{\int \exp \left( \nu(\theta') \sum_{j=1-m}^{T+m} h_2(u_j, z_j, \theta^0) \right) P_0(\text{d}z_{1-m}^{T+m}, \text{d}u_{1-m}^{T+m}, \text{d}z_{T+1}^{T+} | Z_{T}^T = z^T_1) \int \exp \left( \nu(\theta') \sum_{j=1-m}^{T+m} h_2(u_j, z_j, \theta^0) \right) P_0(\text{d}z_{1-m}^{T+m}, \text{d}u_{1-m}^{T+m})}{\int \exp \left( \nu(\theta') \sum_{j=1-m}^{T+m} h_2(u_j, z_j, \theta^0) \right) P_0(\text{d}z_{1-m}^{T+m}, \text{d}u_{1-m}^{T+m})}.
\]

Now in Proposition 1.3.4 I have defined a joint density for \((U_t, Z_t)^T_{t=0}\) with respect to the true DGP.
I write the marginal density of the observables by integrating out the latent variables as:

\[
\prod_{t=1}^{T} \frac{\exp(\nu_{t-1}(\theta)' h_{2,t-1}(z_t, u_t, \theta^0))}{\exp(\nu_{t-1}(\theta)' h_{2,t-1}(z_t, u_t, \theta^0)) P_0(dz_t|I_{t-1})}.
\]

Theorems 1.3.1 and 1.3.2 below deduce semiparametric efficiency bounds for models (1.1) and (1.2) by computing the Fisher information matrices for these marginal likelihoods.

**Theorem 1.3.1.** Suppose Assumptions 1.3.1 and 1.3.2 hold, a semiparametric asymptotic efficiency bound for model (1.1) is given by:

\[
\left\{ E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) | \theta = \theta^0 \right] \right\}' \Sigma_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) | \theta = \theta^0 \right] \right\}^{-1}.
\]

(1.20)

**Theorem 1.3.2.** Suppose Assumptions 1.3.3 and 1.3.4 hold, a semiparametric asymptotic efficiency bound for model (1.2) is given by:

\[
E \left[ E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right] ' \right] Var \left( h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right)^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right] \right]^{-1}.
\]

(1.21)

Notice that as this point there is no guarantee that the efficiency bounds put forth in Theorems 1.3.1 and 1.3.2 are feasible. However, chapter 2 describes feasible estimation procedures that attain these bounds.

### 1.3.5 Feasible efficiency bounds

To conclude the discussion of efficiency bounds for models (1.1) and (1.2) I observe that it is not obvious that the bounds given by (1.20) and (1.21) are feasible. A naive first effort to attain these bounds would be to make use of just identified moment restrictions with the form

\[
E \left[ E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta^0) \right] ' \right] Var \left( h_2(U_t, Z_t, \theta^0) \right)^{-1} h_2(U_t, Z_t, \theta) = 0
\]

\[
E \left[ E_0 \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right] \right] Var \left( h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right)^{-1} h_{2,t-1}(U_t, Z_t, \theta) = 0.
\]
and the corresponding estimating equations:

\[
E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta^0) \right]' \operatorname{Var}(h_2(U_t, Z_t, \theta^0))^{-1} \frac{1}{T} \sum_{t=1}^{T} h_2(U_t, Z_t, \theta) = 0
\]

\[
\frac{1}{T} \sum_{t=1}^{T} E_0 \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right] \operatorname{Var}(h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1})^{-1} h_{2,t-1}(U_t, Z_t, \theta) = 0.
\]

for models (1.1) and (1.2) respectively. As it is clear that these moment restrictions would attain the efficiency bounds of (1.20) and (1.21), it is appropriate to note at this point that a set of optimal instruments, \( Z_{t-1} \) in (1.3), is given by:

\[
Z_{t-1}^{\text{optimal}} = E_0 \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta) | I_{t-1} \right] \operatorname{Var}(h_{2,t-1}(U_t, Z_t, \theta) | I_{t-1})^{-1} B_{2, t}^{-1}.
\]

Unfortunately, direct implementation of these estimating equations and optimal instruments is not feasible both because knowledge of the functions \( h_2, h_{2,t-1} \) is not available without knowledge of \( \Sigma_E, \Sigma_{E,t-1} \) and because the functions \( h_2, h_{2,t-1} \) depend in general on the latent variables when \( \theta \neq \theta^0 \).

However, the following propositions demonstrate that the bounds given by (1.20) and (1.21) are equal to bounds which were obtained intuitively in (1.11) and (1.16). This motivates an appropriate method for attaining the bounds. Specifically, it suggests using an estimation procedure that approximates the least favorable distributions which appeared in (1.11) and (1.16) using a data driven method. As I will show in chapter 2, the ELVIS approach of Schennach (2014) does exactly this.

**Proposition 1.3.5.**

\[
\Gamma_V (\Sigma_V L^R)^{+} \Gamma_V = E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) | \theta = \theta^0 \right]' \Sigma_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) | \theta = \theta^0 \right]
\]

**Proposition 1.3.6.**

\[
E[\Gamma_{t-1}' (\Sigma_{V,t-1} L)^{+} \Gamma_{t-1}] = E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(z_t, u_t, \theta^0) | I_{t-1} \right]' \operatorname{Var}(h_{2,t-1}(z_t, u_t, \theta^0) | I_{t-1})^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(z_t, u_t, \theta^0) | I_{t-1} \right]
\]
1.4 Conclusion

In this chapter I have examined the issue of semiparametric efficiency in models that contain latent variables. The models are defined in terms of some moment restrictions either conditional or unconditional. I have characterized the semiparametric efficiency bounds for these models while allowing for the data generating process to be serially dependent. Further theoretical development would extend the results of this chapter to settings with panel data or data with spatial dependence.

1.5 Appendix: Proofs

1.5.1 Proofs of Propositions

Proof of Proposition 1.3.1:

Proof. The matrix $\Sigma_E$ is a symmetric matrix of rank $r_E$. Therefore, its spectral decomposition can be written as:

$$
\Sigma_E = B' \begin{pmatrix}
\Lambda & 0 \\
0 & 0
\end{pmatrix} B
$$

with $B$ an orthogonal matrix and $\Lambda$ a diagonal matrix of dimension $r_E$ with positive diagonal elements. I define:

$$
g^*(U, Z, \theta) = Bg(U, Z, \theta) = \begin{pmatrix}
g_1^*(U, Z, \theta) \\
h_2(U, Z, \theta)
\end{pmatrix}
$$

Thus:

$$
E[Var(g^*(U, Z, \theta)|Z)] = B\Sigma_E B' = \begin{pmatrix}
\Lambda & 0 \\
0 & 0
\end{pmatrix}
$$

\qed
Proof of Proposition 1.3.2

Proof. The matrix $\Sigma_{E}^{t-1}$ is a symmetric matrix of rank $r_{E}^{t-1}$. Therefore, its spectral decomposition can be written as:

$$
\Sigma_{E}^{t-1} = B^{t-1} \Lambda^{t-1} B^{t-1}
$$

with $B^{t-1}$ an orthogonal matrix and $\Lambda^{t-1}$ a diagonal matrix of dimension $r_{E}^{t-1}$ with positive diagonal elements. As the matrix $\Sigma_{E}^{t-1}$ is $I_{t-1}$-measurable the matrix $B^{t-1}$ is $I_{t-1}$-measurable. I define:

$$
g_{t-1}^{*}(U_{t}, Z_{t}, \theta) = B^{t-1} g(U_{t}, Z_{t}, \theta) = \begin{pmatrix} g_{1, t-1}^{*}(U_{t}, Z_{t}, \theta) \\ g_{2, t-1}^{*}(U_{t}, Z_{t}, \theta) \end{pmatrix}
$$

Thus:

$$
E[\text{Var}(g_{t-1}^{*}(U_{t}, Z_{t}, \theta) | Z_{t}, I_{t-1})] = B^{t-1} \Sigma_{E}^{t-1} B^{t-1} = \begin{pmatrix} \Lambda^{t-1} & 0 \\ 0 & 0 \end{pmatrix}
$$

Proof of Proposition 1.3.3:

Proof. Rather than showing that $\lambda(\theta)$ and $\nu(\theta)$ can be defined so that

$$
\tilde{E}_{\theta}[g(U_{t}, Z_{t}, \theta)] = 0
$$

for all $\theta$ in a neighborhood of $\theta^{0}$, I will instead show the equivalent condition that:

$$
\tilde{E}_{\theta}[h^{*}(U_{t}, Z_{t}, \theta)] = 0
$$

where

$$
h^{*}(U_{t}, Z_{t}, \theta) = N h(U_{t}, Z_{t}, \theta)
$$

with

$$
N = \begin{pmatrix} I & -C(\Sigma_{22}^{m})^{-1} \\ 0 & I \end{pmatrix}
$$

and

$$
C = \text{Cov} \left( h_{1}(U_{t}, Z_{t}, \theta^{0}), \sum_{s=t-m}^{t+m} h_{2}(U_{s}, Z_{s}, \theta^{0}) \right)
$$
Define the following notations

\[ f_{u|z}(u_t : \lambda) = \frac{\int \exp(\sum_{j=t-m}^{t+m} \lambda h_1(u_j, z_j, \theta^0)) P_0(du_{t-m}^{-1}du_{t+1}^{t+m}|Z_t = z_t, U_t = u_t)}{\int \exp(\sum_{j=t-m}^{t+m} \lambda h_1(u_j, z_j, \theta^0)) P_0(du_{t-m}^{k-m}|Z_k = z_k)} \]

\[ f_z(z_t : \nu) = \frac{\exp(\nu h_2(u_t, z_t, \theta^0))}{\int \exp(\nu h_2(u_t, z_t, \theta^0)) P_0(dz_t)} \]

and the following function of \( \theta, \nu \) and \( \lambda \):

\[ H(\theta, \nu, \lambda) = \begin{pmatrix} H_1(\theta, \nu, \lambda) \\ H_2(\theta, \nu, \lambda) \end{pmatrix} = \begin{pmatrix} \int h_1^*(u_t, z_t, \theta)f_{u|z}(u_t : \lambda)f_z(z_t : \nu)P_{0,(U,Z)}(du_t, dz_t) \\ \int h_2(u_t, z_t, \theta)f_{u|z}(u_t : \lambda)f_z(z_t : \nu)P_{0,(U,Z)}(du_t, dz_t) \end{pmatrix} \]

To apply the implicit function theorem about \( (\theta, \nu, \lambda) = (\theta^0, 0, 0) \) I compute:

\[ \frac{\partial H(\theta^0, 0, 0)}{\partial \lambda} |_{\theta=\theta^0, \nu=0, \lambda=0} = E \left[ Cov_0 \left( h_1(U_t, Z_t, \theta^0), h_1(U_t, Z_t, \theta^0)|Z_t \right) \right] = \Lambda \]

\[ \frac{\partial H(\theta^0, 0, 0)}{\partial \nu} |_{\theta=\theta^0, \nu=0, \lambda=0} = Cov_0 \left( h_1(U_t, Z_t, \theta^0), \sum_{s=t-m}^{t+m} h_2(U_s, Z_s, \theta^0) \right) = 0 \]

\[ \frac{\partial H(\theta^0, 0, 0)}{\partial \lambda} |_{\theta=\theta^0, \nu=0, \lambda=0} = Cov_0 \left( h_2(U_t, Z_t, \theta^0), h_1(U_t, Z_t, \theta^0)|Z_t \right) = 0 \]

\[ \frac{\partial H(\theta^0, 0, 0)}{\partial \nu} |_{\theta=\theta^0, \nu=0, \lambda=0} = Cov_0 \left( h_2(U_t, Z_t, \theta^0), \sum_{s=t-m}^{t+m} h_2(U_s, Z_s, \theta^0) \right) = \sum_{m}^{\Lambda} \]
Combining these results:

\[
\frac{\partial H(\theta, \nu, \lambda)}{\partial (\lambda, \nu)} \bigg|_{\theta = \theta^0, \nu = 0, \lambda = 0} = \begin{pmatrix}
\Lambda & 0 \\
0 & \Sigma_{22}^{-m}
\end{pmatrix}
\]

which is positive definite and therefore nonsingular. So by the implicit function theorem there exist functions \(\nu(\cdot), \lambda(\cdot)\) such that for all \(\theta\) in a neighborhood of \(\theta^0\)

\[
H(\theta, \nu(\theta), \lambda(\theta)) = 0
\]

with:

\[
\nu(\theta^0) = 0 \\
\lambda(\theta^0) = 0 \\
\frac{\partial}{\partial \theta} \nu'(\theta^0) = -E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) \big|_{\theta = \theta^0} \right]' (\Sigma_{22}^{-m})^{-1} \\
\frac{\partial}{\partial \theta} \lambda'(\theta^0) = -E \left[ \frac{\partial}{\partial \theta} h_1(U_t, Z_t, \theta) \big|_{\theta = \theta^0} \right]' \Lambda^{-1}
\]

Proof of Proposition 1.3.4

Proof. For any history, \(I_{t-1}\), I define the following function of \(\theta, \nu_{t-1}\) and \(\lambda_{t-1}\):

\[
H_{t-1}(\theta, \nu_{t-1}, \lambda_{t-1}) = \begin{pmatrix}
H_{1,t-1}(\theta, \nu_{t-1}, \lambda_{t-1}) \\
H_{2,t-1}(\theta, \nu_{t-1}, \lambda_{t-1})
\end{pmatrix}
\]

\[
= \left( \begin{array}{c}
\int h_{1,t-1}(z_t, u_t, \theta) \frac{\exp(\lambda' h_{1,t-1}(z_t, u_t, \theta^0))}{\int \exp(\lambda' h_{1,t-1}(z_t, u_t, \theta^0)) P_0(du_t | I_{t-1})} P_0(du_t, dz_t | I_{t-1}) \\
\int h_{2,t-1}(z_t, u_t, \theta) \frac{\exp(\lambda' h_{2,t-1}(z_t, u_t, \theta^0))}{\int \exp(\lambda' h_{2,t-1}(z_t, u_t, \theta^0)) P_0(du_t | I_{t-1})} P_0(du_t, dz_t | I_{t-1})
\end{array} \right)
\]

\[
\int h_{1,t-1}(z_t, u_t, \theta) \frac{\exp(\lambda' h_{1,t-1}(z_t, u_t, \theta^0))}{\int \exp(\lambda' h_{1,t-1}(z_t, u_t, \theta^0)) P_0(du_t | I_{t-1})} P_0(du_t, dz_t | I_{t-1})
\]

\[
\int h_{2,t-1}(z_t, u_t, \theta) \frac{\exp(\lambda' h_{2,t-1}(z_t, u_t, \theta^0))}{\int \exp(\lambda' h_{2,t-1}(z_t, u_t, \theta^0)) P_0(du_t | I_{t-1})} P_0(du_t, dz_t | I_{t-1})
\]
To apply the implicit function theorem about \((\theta, \nu_{t-1}, \lambda_{t-1}) = (\theta^0, 0, 0)\) I compute:

\[
H_{t-1}(\theta^0, 0, 0) = 0
\]

\[
\frac{\partial H_{1,t-1}(\theta, \nu_{t-1}, \lambda_{t-1})}{\partial \lambda_{t-1}}|_{\theta=\theta^0, \nu_{t-1}=0, \lambda_{t-1}=0} = E \left[ \text{Var} \left( h_{1,t-1}(Z_t, U_t, \theta^0) | I_{t-1} \right) \right]
\]

\[
\frac{\partial H_{2,t-1}(\theta, \nu_{t-1}, \lambda_{t-1})}{\partial \nu_{t-1}}|_{\theta=\theta^0, \nu_{t-1}=0, \lambda_{t-1}=0} = \text{Cov} \left( h_{2,t-1}(Z_t, U_t, \theta^0) | I_{t-1} \right)
\]

Combining these results:

\[
\frac{\partial H(\theta, \nu, \lambda)}{\partial(\lambda, \nu)}|_{\theta=\theta^0, \nu=0, \lambda=0} = \begin{pmatrix} \Lambda_{t-1} & 0 \\ 0 & \Sigma_{22}^{t-1} \end{pmatrix}
\]

which is positive definite and therefore nonsingular. So by the implicit function theorem there exist functions \(\nu_{t-1}(\cdot), \lambda_{t-1}(\cdot)\) such that for all \(\theta\) in a neighborhood of \(\theta^0\)

\[
H_{t-1}(\theta, \nu_{t-1}(\theta), \lambda_{t-1}(\theta)) = 0
\]

with:

\[
\nu_{t-1}(\theta^0) = 0
\]

\[
\lambda_{t-1}(\theta^0) = 0
\]

\[
\frac{\partial}{\partial \theta} \nu'_{t-1}(\theta^0) = -E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta) |_{\theta=\theta^0} | I_{t-1} \right]' (\Sigma_{22}^{t-1})^{-1}
\]

\[
\frac{\partial}{\partial \theta} \lambda'_{t-1}(\theta^0) = -E \left[ \frac{\partial}{\partial \theta} h_{1,t-1}(U_t, Z_t, \theta) |_{\theta=\theta^0} | I_{t-1} \right]' (\Lambda_{t-1})^{-1}
\]
Proof of Proposition 1.3.5:

The proof of Proposition 1.3.5 is based on lemmata 1.5.1, 1.5.2, and 1.5.3 below. The notations $C$ and $N$ are defined in the proof of Proposition 1.3.3.

**Lemma 1.5.1.**

\[ \Gamma'_V(S^L VR) + \Gamma_V = E \left[ \frac{\partial}{\partial \theta} \tilde{g}_0^*(Z, \theta^0) \right] \left\{ \text{Var}[E(g^*(U, Z, \theta^0))] \right\}^+ E \left[ \frac{\partial}{\partial \theta} \tilde{g}_0^*(Z, \theta^0) \right] \]

**Lemma 1.5.2.**

\[ E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{1,0}^*(Z, \theta^0) \right] = \left[ \Sigma_{12}^{-1} + C(\Sigma_{22}^m)^{-1} \right] E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{2,0}^*(Z, \theta^0) \right] \]

**Lemma 1.5.3.**

\[ M^{-1} \{ \text{Var}[E(g^*(U, Z, \theta^0)|Z]) \}^+ M^{-1} = \begin{pmatrix} F_1 & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \]

\[ F_1 = [\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}]^+ \]

**Proof.** From Lemma 1.5.1:

\[ \Gamma'_V(S^L VR) + \Gamma_V = E \left[ \frac{\partial}{\partial \theta} \tilde{g}_0^*(Z, \theta^0) \right] \left\{ \text{Var}[E(g^*(U, Z, \theta^0))] \right\}^+ M^{-1} \{ \text{Var}[E(g^*(U, Z, \theta^0)|Z]) \} + M^{-1} NME \left[ \frac{\partial}{\partial \theta} \tilde{g}_0^*(Z, \theta^0) \right] \]

and from Lemma 1.5.2:

\[ NME \left[ \frac{\partial}{\partial \theta} \tilde{g}_0^*(Z, \theta^0) \right] = \begin{pmatrix} 0 \\ E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{2,0}^*(Z, \theta^0) \right] \end{pmatrix} = \begin{pmatrix} 0 \\ E \left[ \frac{\partial}{\partial \theta} g^*(U, Z, \theta^0) \right] \end{pmatrix} \]

Now from Lemma 1.5.3:

\[ N^{-1} M^{-1} \{ \text{Var}[E(g^*(U, Z, \theta^0)|Z]) \}^+ M^{-1} N^{-1} = N^{-1} \begin{pmatrix} F_1 & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} N^{-1} = \begin{pmatrix} F_1 & F_1 C(\Sigma_{22}^m)^{-1} \\ (F_1 C(\Sigma_{22}^m)^{-1})' & \Sigma_{22}^{-1} \end{pmatrix} \]
Proof of Proposition 1.3.6

The proof of Proposition 1.3.6 is based on lemmata 1.5.4, 1.5.5, and 1.5.6 below. The equalities below are understood to hold with probability one.

Lemma 1.5.4.

\[ \Gamma_{t-1}(\Sigma_{V}^{l,p})^{+} \Gamma_{t-1} = E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{1,t-1}^{\ast}(U_{t},Z_{t},\theta^{0}) \right]^{\prime} \left( F_{1} F_{1} C(\Sigma_{22}^{m})^{-1} \right) \left( (F_{1} C(\Sigma_{22}^{m})^{-1})^{\prime} \Sigma_{22}^{-1} \right) E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{2}^{*}(U_{t},Z_{t},\theta^{0}) \right] \]

Lemma 1.5.5.

\[ E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{1,t-1}^{\ast}(Z_{t},\theta^{0})|I_{t-1} \right] = \Sigma_{12}^{-1}(\Sigma_{22}^{-1})^{-1} E \left[ \frac{\partial}{\partial \theta} \tilde{g}_{2,t-1}^{\ast}(Z_{t},\theta^{0})|I_{t-1} \right] \]

Lemma 1.5.6.

\[ (M^{t-1})^{-1} \left( \text{Var} E(g^{*}(U_{t},Z_{t},\theta^{0})|Z_{t},I_{t-1}) \right) + (M^{t-1})^{-1} = \left( \begin{array}{cc} F_{1}^{-1} & 0 \\ 0 & (\Sigma_{22}^{-1})^{-1} \end{array} \right) \]

\[ F_{1}^{-1} = \Sigma_{11}^{-1} - \Sigma_{12}^{-1} (\Sigma_{22}^{-1})^{-1} \Sigma_{21}^{-1} \]

Proof. I will actually prove the stronger result:

\[ \Gamma_{t-1}(\Sigma_{V}^{l,p})^{+} \Gamma_{t-1} = E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_{t},Z_{t},\theta^{0})|I_{t-1} \right] \left( \text{Var} h_{2,t-1}(U_{t},Z_{t},\theta^{0})|I_{t-1} \right)^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_{t},Z_{t},\theta^{0})|I_{t-1} \right] \]
From Lemma 1.5.4:

\[ \Gamma'_{t-1}(\Sigma_{V_{t-1}})^{-1} + \Gamma_{t-1} = E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, \theta}(Z_t, \theta^0)|I_{t-1} \right]' \left( M^{t-1} \right)^{-1} \left\{ \right. \]

\[ \left. \left[ \text{Var}[E(g_{t-1}((U_t, Z_t, \theta^0))|I_t)|I_{t-1}] \right] + \left( M^{t-1} \right)^{-1} M^{t-1} E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, \theta}(Z_t, \theta^0)|I_{t-1} \right] \right\} \]

and from Lemma 1.5.5:

\[ M^{t-1} E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, \theta}(Z_t, \theta^0)|I_{t-1} \right] = \begin{pmatrix} 0 \\ E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right] \end{pmatrix} = \begin{pmatrix} 0 \\ E\left[ \frac{\partial}{\partial \theta} g_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right] \end{pmatrix} \]

Now from Lemma 1.5.6:

\[ \Gamma'_{t-1}(\Sigma_{V_{t-1}})^{-1} + \Gamma_{t-1} = \begin{pmatrix} 0, E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right] \end{pmatrix}' \left( F_1 \begin{pmatrix} 0 \\ (\Sigma_{22}^{-1})^{-1} \end{pmatrix} \right) \begin{pmatrix} 0 \\ E\left[ \frac{\partial}{\partial \theta} g_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right] \end{pmatrix} \]

\[ = E\left[ \frac{\partial}{\partial \theta} \tilde{g}_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right]' (\Sigma_{22}^{-1})^{-1} E\left[ \frac{\partial}{\partial \theta} g_{t-1, 2,(U_t, Z_t, \theta^0)}|I_{t-1} \right] \]

\[ \square \]

1.5.2 Proofs of Lemmata

**Proof of lemma 1.2.1**

*Proof.* From the second and third components of (1.4) I deduce:

\[ E[y_t x_{t-1}^*] = \beta E[x_t^* x_{t-1}^*] \]
\[ E[y_{t-1} x_{t}^*] = \beta E[x_t^* x_{t-1}^*] \]

Note that:

\[ (y_t - \beta x_t^*)(y_{t-1} - \beta x_{t-1}^*) = y_t y_{t-1} - \beta(x_t^* y_{t-1} + y_t x_{t-1}^*) + \beta^2 x_t^* x_{t-1}^* \]

Thus,

\[ E[(y_t - \beta x_t^*)(y_{t-1} - \beta x_{t-1}^*)] = E[y_t y_{t-1}] - \beta^2 E[x_t^* x_{t-1}^*] \]
So that the seventh component of (1.4) implies:

\[ E[y_t y_{t-1}] = \beta^2 E[x^*_t x^*_t - 1] \]

Summing together the second and sixth components of (1.4) yields:

\[ E[y_t x_{t-1}] = \beta E[x^*_t x_{t-1}] \]

and from the twelfth component of (1.4):

\[ \beta E[x^*_t x_{t-1}] = \beta E[x_t x_{t-1}] \]

Therefore when \( \beta \neq 0 \):

\[ \beta = \frac{E[y_t y_{t-1}]}{E[y_t x_{t-1}]} \]

And from the first and fourth components of (1.4):

\[ \beta = 0 \iff E[y_t x_t] = 0 \]

\[ \square \]

**Proof of lemma 1.2.2**

*Proof.* Note that:

\[(y_t - B f_t)(y_t - B f_t)' = y_t y_t' - B f_t y_t' - Y_t f_t' B' + B f_t f_t' B'\]

From the fourth and fifth components of (1.6) deduce:

\[ E(f_t f_t') = I \]

\[ E(Y_t f_t') = B \]
Thus:

\[ E[(y_t - Bf_t)(y_t - Bf_t)'] = E[y_ty'_t] - BB' \]

So that the third component of (1.6) implies:

\[ E[y_ty'_t] = BB' + \Phi \]

The left side is an identified quantity and contains \( \frac{n(n+1)}{2} \) unique components. The right side contains \( nq + n \) parameters to be identified.

\[ \square \]

**Proof of lemma 1.2.3**

*Proof.* Following the same computations as in the proof of lemma 1.2.2 I obtain:

\[ E[y_ty'_t] = BB' + \Phi \]

As in lemma 1.2.2, this equation identifies \( B \) and \( diag(\Phi) \).

I now turn our attention to the identification of the matrix \( A \). Note that:

\[ (y_t - Bf_t)(f_t - Af_{t-1})' = y_tf'_t - Bf_tf'_t - y_tf'_{t-1}A' + Bf_tf'_{t-1}A' \]

From the fourth and sixth components of (1.7) I deduce that:

\[ E(y_tf'_t) = B \]

and from the fifth and seventh components of (1.7):

\[ E(f_tf'_{t-1}) = A \]

Thus:

\[ E[(y_t - Bf_t)(f_t - Af_{t-1})'] = B - B - E[y_tf'_{t-1}]A' + BAA' \]
So that from the eighth component of (1.7) I obtain:

\[ E[y_t f'_{t-1}] = BA \]

And from the tenth and eleventh components of (1.7):

\[ E[f_t y'_{t-1}] = AB \]

Next note that:

\[(y_t - B f_t)(y_{t-1} - B f_{t-1})' = y_t y'_{t-1} - B f_t y'_{t-1} - y_t f'_{t-1} B' + B f_t f'_{t-1} B'\]

Thus:

\[ E[(y_t - B f_t)(y_{t-1} - B f_{t-1})'] = E[y_t y'_{t-1}] - BAB' - BAB' + BAB' \]

So that the ninth component of (1.7) implies:

\[ E[y_t y'_{t-1}] = BAB' \]

Since \( B \) is of full column rank I can solve for \( A \) as:

\[ A = (B'B)^{-1} B' E[y_t y'_{t-1}] B (B'B)^{-1} \]

Proof of lemma 1.2.4

**Proof.** Following the same computations as in the proof of lemma 1.2.2 I obtain:

\[ E[y_t y'_t] = BB' + \Phi \]

As in lemma 1.2.2, this equation identifies \( B \) and \( \text{diag}(\Phi) \).
From the sixth and seventh component of (1.9) I have that:

\[ E[y_t y_{t-1}'] \{ E[y_{t-1} y_{t-1}'] \}^{-1} = C \]

\[ \square \]

**Proof of Lemma 1.3.1:**

*Proof.* From Proposition 1.3.1:

\[ E \left[ \text{Var} (h_2(U_t, Z_t, \theta^0) | Z_t) \right] = 0 \]

However, for any square-integrable random variable \( Y \):

\[ E[ \text{Var}(Y | Z)] = 0 \]

\[ \Rightarrow \text{Var}(Y) = \text{Var}[E(Y | Z)] \]

\[ \Rightarrow \text{Var}[Y - E(Y | Z)] = \text{Var}(Y) - \text{Var}[E(Y | Z)] = 0 \]

\[ \Rightarrow Y =_{as} E[Y | Z] \]

\[ \square \]

**Proof of lemma 1.3.2**

*Proof.* From proposition 1.3.2:

\[ E[\text{Var}(h_2(U_t, Z_t, \theta^0) | I_t) | I_{t-1}] = 0 \]

\[ \Rightarrow E[\text{Var}(h_2(U_t, Z_t, \theta^0) | I_t)] = 0 \]

so that the argument of lemma 1.3.1 applies.

\[ \square \]
Proof of Lemma 1.5.1:

Proof. Since $B$ is an orthogonal matrix:

$$
\Gamma_V'(\Sigma_V^{LR})^+\Gamma_V = (B\Gamma_V)'B(\Sigma_V^{LR})^+B'(B\Gamma_V)
$$

Letting $A = B(\Sigma_V^{LR})^+B'$ the following is true:

$$
A^+ = B\Sigma_V^{LR}B' 
\quad \text{(1.22)}
$$

$$
A = [B\Sigma_V^{LR}B']^+ 
\quad \text{(1.23)}
$$

To see (1.22), recall that the Moore-Penrose inverse of $A$ is the unique matrix $A^+$ that satisfies the following equations:

$$
AA^+A = A
$$

$$
A^+AA^+ = A^+
$$

$$
( AA^+ )' = AA^+
$$

$$
( A^+A )' = A^+A
$$

And (1.23) follows from the fact that $(A^+)^+ = A$.

Proof of Lemma 1.5.2:

Proof. I note that:

$$
\tilde{g}_{1,\theta}(z,\theta^0) = \int g_1^*(u, z, \theta) \exp\{\lambda(\theta)'h_1(u, z, \theta^0)\} dP_{0, U|Z}(u|z)
$$

Therefore:

$$
\frac{\partial}{\partial \theta'} \tilde{g}_{1,\theta}(z,\theta^0) = \int \frac{\partial g_1^*(u, z, \theta)}{\partial \theta'} \exp\{\lambda(\theta)'h_1(u, z, \theta^0)\} dP_{0, U|Z}(u|z)
$$

$$
+ \int g_1^*(u, z, \theta)h_1(u, z, \theta^0) \frac{\partial \lambda(\theta)}{\partial \theta'} \exp\{\lambda(\theta)'h_1(u, z, \theta^0)\} dP_{0, U|Z}(u|z)
$$
Hence:

\[ E \left( \frac{\partial}{\partial \theta} \tilde{g}_1^*(\mathbf{Z}, \theta^0) \right) = E \left[ \frac{\partial g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] + E \left[ \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_1(\mathbf{U}, \mathbf{Z}, \theta^0) \right) \right] \frac{\partial \lambda(\theta^0)}{\partial \theta} \]

I compute:

\[ \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_1(\mathbf{U}, \mathbf{Z}, \theta^0) \right) = \text{Var} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0) \right) - \Sigma_{12} \Sigma_{22}^{-1} \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_2(\mathbf{U}, \mathbf{Z}, \theta^0) \right) \]

\[ \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_1(\mathbf{U}, \mathbf{Z}, \theta^0) | \mathbf{Z} \right) = \text{Var} \left[ E \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0) | \mathbf{Z} \right) \right] - \Sigma_{12} \Sigma_{22}^{-1} \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_2(\mathbf{U}, \mathbf{Z}, \theta^0) \right) \]

So that:

\[ E \left[ \text{Cov} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0), h_1(\mathbf{U}, \mathbf{Z}, \theta^0) | \mathbf{Z} \right) \right] = \text{Var} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0) \right) - \text{Var} \left[ E \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0) | \mathbf{Z} \right) \right] \]

\[ = E \left[ \text{Var} \left( g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0) \right) \right] = \Lambda \]

Finally,

\[ E \left( \frac{\partial}{\partial \theta} \tilde{g}_1^*(\mathbf{Z}, \theta^0) \right) = E \left[ \frac{\partial g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] + \Lambda \frac{\partial \lambda(\theta^0)}{\partial \theta} \]

\[ = E \left[ \frac{\partial g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] - E \left[ \frac{\partial h_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] \]

where the last equality comes from Proposition 1.3.3. The result then follows from the definition of \( h_1^* \):

\[ E \left[ \frac{\partial g_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] - E \left[ \frac{\partial h_1^*(\mathbf{U}, \mathbf{Z}, \theta^0)}{\partial \theta} \right] = [\Sigma_{12} \Sigma_{22}^{-1} + C(\Sigma_{22}^{\text{m}})^{-1}] E \left[ \frac{\partial \tilde{g}_2^* \theta}{\partial \theta} \right] \]
Proof of Lemma 1.5.3:

Proof. Recall that:

\[ B\Sigma_{V}^{LR}B' = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \]

where \( \Sigma_{22} \) is nonsingular by A3. To compute the Moore-Penrose inverse of \( B\Sigma_{V}^{LR}B' \) I use a formula given by Phillips (1992):

\[
(B\Sigma_{V}^{LR}B')^+ = \begin{pmatrix} F_1 & -F_1\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma'_{12}F_1 & F_2 \end{pmatrix}
\]

\[ F_2 = [\Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{+}\Sigma_{12}]^{-1} \]

Thus:

\[
(B\Sigma_{V}^{LR}B')^+ M^{-1} = (B\Sigma_{V}^{LR}B')^+ \begin{pmatrix} I & \Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix}
\]

\[ = \begin{pmatrix} F_1 & 0 \\ -\Sigma_{22}^{-1}\Sigma'_{12}F_1 & A \end{pmatrix}
\]

\[ A = -\Sigma_{22}^{-1}\Sigma'_{12}F_1\Sigma_{12}\Sigma_{22}^{-1} + F_2 \]

and:

\[
M^{-1}(B\Sigma_{V}^{LR}B')^+ M^{-1} = \begin{pmatrix} I & 0 \\ \Sigma_{22}^{-1}\Sigma'_{12} & I \end{pmatrix} \begin{pmatrix} F_1 & 0 \\ -\Sigma_{22}^{-1}\Sigma'_{12}F_1 & A \end{pmatrix}
\]

\[ = \begin{pmatrix} F_1 & 0 \\ 0 & A \end{pmatrix} \]
From Phillips (1992)

\[ F_1 = \Sigma_{11}^+ + \Sigma_{11}^+ \Sigma_{12} F_2 \Sigma_{12}^+ \Sigma_{11}^+ \]

which implies:

\[
F_1 \Sigma_{12} \Sigma_{22}^{-1} = [\Sigma_{11}^+ + \Sigma_{11}^+ \Sigma_{12} F_2 \Sigma_{12}^+ \Sigma_{11}^+] \Sigma_{12} \Sigma_{22}^{-1} \\
= \Sigma_{11}^+ \Sigma_{12} \Sigma_{22}^{-1} + \Sigma_{11}^+ \Sigma_{12} F_2 \Sigma_{12}^+ \Sigma_{12} \Sigma_{22}^{-1} \\
= \Sigma_{11}^+ \Sigma_{12} \Sigma_{22}^{-1} \{I + [I - \Sigma_{12} \Sigma_{12} \Sigma_{22}^{-1}]^{-1} \Sigma_{12} \Sigma_{12} \Sigma_{22}^{-1}\} \\
= \Sigma_{11}^+ \Sigma_{12} \Sigma_{22}^{-1} [I - \Sigma_{12} \Sigma_{12} \Sigma_{22}^{-1}]^{-1} \\
= \Sigma_{11}^+ \Sigma_{12} F_2
\]

With this result I compute the matrix \( A \) as:

\[
A = -\Sigma_{22}^{-1} \Sigma_{12} F_1 \Sigma_{12} \Sigma_{22}^{-1} + F_2 \\
= -\Sigma_{22}^{-1} \Sigma_{12} \Sigma_{11}^+ \Sigma_{12} F_2 + F_2 \\
= [I - \Sigma_{22}^{-1} \Sigma_{12} \Sigma_{11}^+ \Sigma_{12}] F_2 \\
= [I - \Sigma_{22}^{-1} \Sigma_{12} \Sigma_{11}^+ \Sigma_{12}] [I - \Sigma_{22}^{-1} \Sigma_{12} \Sigma_{11}^+ \Sigma_{12}]^{-1} \Sigma_{22}^{-1} \\
= \Sigma_{22}^{-1}
\]

Hence I obtain the formula given in lemma 1.5.3.

\[
\square
\]

**Proof of Lemma 1.5.4:**

*Proof.* The proof is the same as that of lemma 1.5.1 up to an obvious change in notation.

\[
\square
\]

**Proof of Lemma 1.5.5:**

*Proof.* The proof is the same as that of lemma 1.5.2 up to an obvious change in notation.

\[
\square
\]

**Proof of Lemma 1.5.6:**

*Proof.* The proof is the same as that of lemma 1.5.3 up to an obvious change in notation.

\[
\square
\]
1.5.3 Proofs of Theorems

Proof of Theorem 1.3.1:

Proof. The score vector for $\theta = \theta^0$ is given by

$$s_T(\theta^0) = \frac{\partial}{\partial \theta} \nu'(\theta^0) \int \sum_{j=-m+1}^{T+m} h_2(u_j, z_j, \theta^0) P_0(dz_{s-m}^{s-1}, dz_{s-m}^{s+m}, du_{s-m}^{s-m}|Z_t^T = z_t^T).$$

Then, the corresponding asymptotic Fisher information matrix is:

$$E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) |_{\theta = \theta^0} \right]' (\Sigma_{22}^m)^{-1} \Sigma_{22}^m (\Sigma_{22}^m)^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) |_{\theta = \theta^0} \right].$$

As $m \to \infty$ the Fisher information becomes arbitrarily close to:

$$E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) |_{\theta = \theta^0} \right]' \Sigma_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta) |_{\theta = \theta^0} \right].$$

Assumptions 1.3.1 and 1.3.2 imply that the previous display is invertible.

Proof of Theorem 1.3.2:

Proof. The score vector for $\theta = \theta^0$ is given by:

$$s_T(\theta^0) = \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \nu_{t-1}((\theta^0)' h_2^{-1}(z_t, u_t, \theta^0)).$$

Then the asymptotic Fisher information matrix is given by:

$$E \left[ E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right]' \right] Var(h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1})^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | I_{t-1} \right].$$

By Assumption 1.3.4 the previous display is invertible.
Chapter 2

Estimation with Latent Variables

2.1 Introduction

This chapter considers methods for the efficient estimation of the models (1.1) and (1.2) introduced in chapter 1.

My approach to estimation of these models is to adapt the approach of Schennach (2014) for accommodating latent variables in partially identified moment restriction models. Schennach’s approach, called entropic latent variable integration via simulation (ELVIS), eliminates latent variables from the moment function $g$ by integrating against a cleverly chosen conditional probability distribution. After the latent variables are eliminated model (1.1) can be estimated using a generalized method of moments (GMM) or generalized empirical likelihood (GEL) type estimator. Model (1.2) is estimated using a local GEL type estimator in the spirit of Kitamura et al. (2004), Smith (2007), Antoine et al. (2007), and Gospodinov and Otsu (2012).

Within the measurement error framework I conduct Monte Carlo simulations and estimate the leverage ratio on expected consumption growth using the long-run risk model of Bansal and Yaron (2004). Both the Monte Carlo simulations and the empirical example suggest that my estimator performs favorably to extant methods.

Although the efficient estimation of model (1.1) has been examined in the case of cross sectional data by Bedard and Renault (2018), the results presented here are the first to consider serially dependent data generating processes.

I make the connection between the work of Schennach (2014) and the theory of efficient point
estimation while allowing for serially dependent data. Furthermore, the one-step estimation procedures I consider fall within the class of estimators that Schennach has shown perform well in partially identified models. This suggests an adaptive procedure for estimation problems where the researcher is unable to verify point identification analytically that does not sacrifice efficiency in the case of point identification.

The two-step efficient estimation procedure for model (1.1) is introduced in section 2.2. Monte Carlo results for the two-step estimator are presented in section 2.3. Estimation results from the long-run risk model of Bansal and Yaron (2004) are discussed in section 2.4. A discussion of one-step estimation for model (1.1) appears in section 2.5. The one-step estimator for model (1.2) is presented in section 2.6. Section 2.7 concludes. All proofs appear in the appendix.

2.2 Two-step estimation in the unconditional model

In this section I describe a feasible estimator of model (1.1) that attains the efficiency bound (1.20). This two-step estimator is similar in spirit to the two-step efficient GMM estimator of Hansen (1982). However, the latent variables included in the moment restrictions must be accounted for. As the parametric model considered in chapter 1 is constructed using the unknown true distribution of \((U_T, Z_T)\), maximum likelihood estimation is not feasible. I instead follow the idea of Schennach (2014) to eliminate latent variables through ELVIS.

The idea is to eliminate latent variables by integration against an exponential tilted family of conditional distributions for the latent variables given the observable variables. I call the estimator discussed in this section the two-step efficient ELVIS estimator. Intuitively, this procedure will compensate for the fact that the true conditional distribution of latents given observables is unknown by introducing a vector of \(H\) nuisance parameters. The estimator of this section was first put forth by Bedard and Renault (2018) for cross sectional applications.

2.2.1 Preliminaries

I now describe the two-step efficient ELVIS estimator. I define the empirical tilted moment conditions as

\[
g_{\theta, \rho}(z, \theta, \gamma) = \frac{\int g(u, z, \theta) \exp(\gamma' g(u, z, \theta)) d\rho(u|z, \theta)}{\int \exp(\gamma' g(u, z, \theta)) d\rho(u|z, \theta)}
\]
where \( \rho(\cdot|z;\theta) \) is a family of conditional probability distributions used as a reference point for the exponential tilting that is indexed by the real vector \( \gamma \in \mathbb{R}^H \) and the parameter of interest \( \theta \). The main result of this section is that extremum estimators obtained by minimizing objective functions of the form

\[
\hat{Q}_T(\theta, \gamma) = \left( \frac{1}{T} \sum_{t=1}^T g_{\theta, \rho}(z_t, \theta, \gamma) \right) \hat{W}_T \left( \frac{1}{T} \sum_{t=1}^T g_{\theta, \rho}(z_t, \theta, \gamma) \right)
\]

(2.1)

when \( \hat{W}_T \to^p W \), a properly chosen positive semidefinite matrix to be described below, achieve the asymptotic efficiency bound for the estimation of \( \theta \). Despite the apparent similarity with the classical GMM objective function the objective function in (2.1) is distinct due to the inclusion of a vector of nuisance parameters \( \gamma \). Notice also that the reference measure \( \rho(\cdot|z;\theta) \) enters the objective function through the function \( g_{\theta, \rho} \). I will show that this reference measure does not influence the numerical value of my estimator using Assumption 2.2.1 below.

In the sequel I will let \( P^Z(\cdot) \) denote the true marginal distribution of a single observation \( Z \). I require the following technical assumptions about \( \rho(\cdot|z;\theta) \) for estimation:

**Assumption 2.2.1.** The user specified reference measure has a support that contains the true support of \( U \) and is such that

\[
\int \log \left( \int \exp(\gamma'g(u, z, \theta))\rho(du|z;\theta) \right) P^Z(dz)
\]

exists and is twice continuously differentiable in \( \gamma \) for all \( \theta \in \Theta \) and \( \gamma \in \mathbb{R}^H \).

Since Schennach (2014) has shown that a reference measure satisfying this assumption always exists, Assumption 2.2.1 is not restrictive. The following proposition summarizes the important fact that, under Assumption 2.2.1, the minimizer \( \hat{\theta} \) of equation (2.1) is numerically independent of the particular reference measure \( \rho(\cdot|z;\theta) \) that has been used.

**Proposition 2.2.1.** Suppose Assumption 2.2.1 holds. Let \( \hat{\theta} = \arg \min_\theta \inf_{\gamma \in \mathbb{R}^H} \hat{Q}_T(\theta, \gamma) \), the value of \( \hat{\theta} \) is independent of the reference measure \( \rho \).

The importance of Proposition 2.2.1 is that I may, when deriving the asymptotic properties of the two-step efficient ELVIS estimator assume that the researcher has chosen a dominating measure \( \rho \) that is computationally convenient and has an intuitive interpretation without any loss of generality.
Assumption 2.2.2. The true conditional probability of $U$ given $Z$, $P_{0,U|Z}(\cdot|z)$, is such that

$$\int \log \left( \int \exp(\gamma' g(u, z, \theta)) P_{0,U|Z}(du|z) \right) P^Z(dz)$$

exists and is twice continuously differentiable in $\gamma$ for all $\theta \in \Theta$ and $\gamma \in \mathbb{R}^H$.

Assumption 2.2.2 says that the true conditional distribution satisfies the conditions of Assumption 2.2.1. This assumption implies that the family of conditional probability measures

$$\frac{dP_{\theta,U|Z}(u_t|Z_t = z_t)}{dP_{0,U|Z}(u_t|Z_t = z_t)} \propto \exp \left( \lambda(\theta)' h_1(u_t, z_t, \theta^0) \right)$$

with $\lambda(\cdot)$ defined in Proposition 1.3.3 is a valid choice of $\rho$. By virtue of Proposition 2.2.1 I may derive the asymptotic properties of the two-step efficient ELVIS estimator assuming, without loss of generality, that the dominating measure used for estimation is given by (2.3).

### 2.2.2 Consistency

I will now give conditions for the consistency of the two-step efficient ELVIS estimator. Assumption 2.2.3 below is the global identification condition for model (1.1) when $U$ is not observed.

Assumption 2.2.3. (Identification) The weighting matrix is positive semi-definite

$$E(g(U, Z, \theta)) = \int \int g(u, z, \theta) P_{0,U|Z}(du|z) P^Z(dz)$$

$$= 0$$

if $\theta^* = \theta^0$ and for all $\theta \neq \theta^0$ there does not exist any conditional probability measure $Q(\cdot|z)$ with support equal to the support of $U$ or a measurable subset of the support of $U$ such that:

$$W \left( \int \int g(u, z, \theta^*) Q(du|z) P^Z(dz) \right) = 0.$$  

(2.4)

Assumption 2.2.3 ensures that no values of $\theta \neq \theta^0$ satisfy the empirical titled moment conditions. If the matrix $W$ is positive definite than the second display can be reduced to:

$$\int \int g(u, z, \theta^*) Q(du|z) P^Z(dz) = 0.$$
The existence of a conditional probability measure \( Q \) that solves the previous equation with \( \theta^* \neq \theta^0 \) means that the parameter is not point identified by the moment restrictions (1.1). As the researcher only observes information about the marginal distribution \( P^Z \), there is no hope to differentiate between \( \theta^0 \) and \( \theta^* \) on the basis of the information available in the moment restrictions which define the semiparametric model. If there exists some linear combination of the moment restrictions that eliminates the latent variables and identifies the parameter vector it is clear that no such \( Q \) can exist. This is the approach taken in chapter 1 to show identification in the examples.

The derivation of asymptotic properties for the estimator \( \hat{\theta} \) is complicated somewhat by the fact that the nuisance parameters \( \gamma \) are not point identified in models where \( \theta \) is identified. I will call any vector of the nuisance parameters that minimizes the probability limit of the sample objective function in equation (2.1) the pseudo-true nuisance parameters. The pseudo-true nuisance parameters are therefore any vector \( \gamma \in \mathbb{R}^H \) that satisfies:

\[
W \left( \int \int g(U, Z, \theta^0)\rho^*(du|z; \gamma, \theta^0)\mu(dz) \right) = 0. \tag{2.5}
\]

Where once again if the matrix \( W \) is positive definite, it may be omitted and \( \rho^*(\cdot|z; \gamma, \theta) \) is the titled reference measure:

\[
\rho^*(\cdot|z; \gamma, \theta) = \frac{\exp(\gamma'g(u, z, \theta))d\rho(u|z, \theta)}{\int \exp(\gamma'g(u, z, \theta))d\rho(u|z, \theta)} \tag{2.6}
\]

One way to observe the lack of uniqueness for \( \gamma \) is by inspection of equation (2.6). If any component of \( g(U, Z, \theta^0) \) is \( Z \)-measurable than the corresponding component of \( \gamma \) is irrelevant to the calculation on the right side of equation (2.6). Although \( \gamma \) is not point identified, the probability measure \( P_\gamma(\cdot|\cdot) = \rho^*(\cdot|\cdot; \gamma, \theta^0)P^Z(\cdot) \) associated with values of \( \gamma \) that satisfy equation (2.5) is unique. This convenient property is summarized in the following lemma.

**Lemma 2.2.1.** For any two vectors \( \gamma_1, \gamma_2 \) that satisfy equation (2.5), the Kullback–Leibler divergence of the associated probability distributions is zero.

In the sequel I use \( G^0 \) to denote the set of \( \gamma \in \mathbb{R}^H \) such that equation (2.5) is satisfied and use \( \gamma^0 \) to denote a generic element of \( G^0 \). This set is dependent on the reference measure \( \rho \). Lemma 2.2.2 shows that \( G^0 \) is convex.
Lemma 2.2.2. The set $G^0$ is convex.

To state the final conditions required for consistency I need the following notations:

$$Q_0(\theta, \gamma) = -E(g_{\theta, \rho}(Z, \theta, \gamma))'WE(g_{\theta, \rho}(Z, \theta, \gamma))$$

$$\hat{Q}(\theta, \gamma) = -\left(\frac{1}{T} \sum_{t=1}^{T} g_{\theta, \rho}(z_t, \theta, \gamma)\right)'\hat{W}\left(\frac{1}{T} \sum_{t=1}^{T} g_{\theta, \rho}(z_t, \theta, \gamma)\right)$$

$$\Delta = \Theta \times \mathbb{R}^H$$

$$\Delta_{0,k} = \left\{ (\theta^*, \gamma^*) \in \Delta : Q_0(\theta^*, \gamma^*) > \sup_{(\theta, \gamma) \in \Delta} Q_0(\theta, \gamma) - \frac{1}{k} \right\}.$$

Assumption 2.2.4 gives the technical conditions required to prove consistency. Since the nuisance parameters are not constrained to take values in a compact set, my approach to consistency is to augment the classical argument given by Wald (1949) for maximum likelihood estimates. For a modern treatment of the Wald consistency proof see, for example, van der Vaart (2000).

Assumption 2.2.4. 1. For any $(\theta, \gamma) \in \Delta$ there exists an open set $V$ such that (i) $(\theta, \gamma) \in V$ and (ii)

$$E\left(\sup_{(\theta^*, \gamma^*) \in V} \|g_{\theta, \rho}(Z, \theta^*, \gamma^*)\|\right) < \infty.$$

2. $(\hat{\theta}_T, \hat{\gamma}_T)$ are such that for all $T$ there exist $(\theta_{0,T}, \gamma_{0,T}) \in \Delta_{0,k}$ satisfying the inequality

$$\hat{Q}(\hat{\theta}_T, \hat{\gamma}_T) \geq \hat{Q}(\theta_{0,T}, \gamma_{0,T}) - \frac{1}{T}.$$

3. For all $(\theta, \gamma) \in \Delta$

$$\text{plim} \hat{Q}(\theta, \gamma) = Q_0(\theta, \gamma).$$

I am now prepared to state the consistency result:

Theorem 2.2.1.

Suppose Assumptions 2.2.1, 2.2.3, and 2.2.4 hold, $g(z, \theta, \gamma)$ is continuous at each $(\theta, \gamma) \in \Theta \times \mathbb{R}^H$ with probability one, $(\theta_T, \gamma_T)_{n=1}^\infty$ are a sequence of point estimates as in Assumption 2.2.4.2, and
\[ P(\|\hat{\theta}\|^2 + \|\hat{\gamma}\|^2 > k^2) \to 0 \text{ for some positive integer } k. \] Then \( \hat{\theta}_T \to^p \theta^0 \) and \( d(G^0, \hat{\gamma}_T) \to^p 0. \)

As a consequence of this result, \( \hat{\theta} \) is consistent and the estimated nuisance parameters become arbitrarily close to values in \( G^0 \) with probability approaching one. As an abuse of terminology, I will refer to the property that \( d(G^0, \hat{\gamma}_T) \to^p 0 \) as consistency for the estimated nuisance parameters. By the almost sure continuity of \( g_{\theta, \rho}(z, \gamma, \theta) \) in \( \gamma \) and \( \theta \) I obtain convergence of the associated densities

\[ \frac{d\rho^*(u|z; \hat{\gamma}_T, \hat{\theta}_T)}{d\rho(u|z; \theta_T)} \to^p \frac{d\rho^*(u|z; \gamma^0, \theta^0)}{d\rho(u|z; \theta^0)} \]

where the right side is uniquely defined by lemma 2.2.1.

### 2.2.3 Asymptotic Normality

In this section, I give the asymptotic normality result for the two-step efficient ELVIS estimator. Theorem 2.2.2 below says that if \( \hat{W}_T \to^P (\Sigma_V^{LR})^+ \) then minimization of (2.1) yields an asymptotically normal and efficient estimator of \( \theta^0 \). However, a method to consistently estimate \( (\Sigma_V^{LR})^+ \) is not at all obvious. Theorem 2.2.2 is presented for expository purposes while a feasible efficient procedure is given by Theorem 2.2.3. In addition to the conditions required for consistency, I require the local identification condition:

**Assumption 2.2.5.** (Local Identification 1) The matrix \( \Gamma'_V (\Sigma_V^{LR})^+ \Gamma_V \) is nonsingular.

Assumption 2.2.5 is an analog of the standard local identification condition in GMM estimation that allows for singular weighting matrices. By virtue of Proposition 1.3.5, Assumption 2.2.5 is implied by Assumptions 1.3.1 and 1.3.2 introduced in chapter 1 to derive an efficiency bound. I am now prepared to state the first asymptotic normality result.

**Theorem 2.2.2.** Suppose that the conditions of Theorem 2.2.1 and Assumptions 2.2.2 and 2.2.5 hold, \( \theta^0 \in \text{interior}(\Theta) \), \( g_{\theta, \rho}(Z, \theta, \gamma) \) is continuously differentiable in a neighborhood \( N \) of \( \theta^0 \), the derivatives of \( g_{\theta, \rho}(z, \theta, \gamma) \) satisfy integrability conditions sufficient for a uniform law of large numbers to apply and \( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \bar{g}_{\theta, \rho}(z_t, \theta^0) \to^d N(0, \Sigma_V^{LR}) \). Then

\[ \sqrt{T}(\hat{\theta} - \theta^0) \to^d N[0, (\Gamma'_V (\Sigma_V^{LR})^+ \Gamma_V)^{-1}]. \]

---

1The distance between a set and a point in \( \mathbb{R}^H \) is understood to mean \( d(A, p) = \inf_{x \in A} \|x - p\| \).
I now consider the case where $\hat{W} \rightarrow_p W$ a positive definite matrix that is not necessarily equal to $(\Sigma_{VR}^{LR})^+$. 

**Assumption 2.2.6.** (Local Identification 2) The matrix $W$ is positive definite and $\Gamma_V$ has full column rank.

Selection of an optimal weighting matrix $W$ is discussed below. The requirement that $\Gamma_V$ has full column rank is implied by Assumption 1.3.1 introduced in chapter 1.

I define the matrices $W_{11}, W_{12}$ and $W_{22}$ through the partitioned matrix equation

$$BW^{-1}B' = \begin{pmatrix} W_{11} & W_{12} \\ W_{12}' & W_{22} \end{pmatrix}$$

where $W_{11}$ is $r_E \times r_E$.

**Theorem 2.2.3.** Suppose that the conditions of Theorem 2.2.1 and Assumptions 2.2.2 and 2.2.6 hold, $\theta^0 \in \text{interior}(\Theta)$, $g_{\theta, \rho}(Z_t, \theta, \gamma)$ is continuously differentiable in a neighborhood $N$ of $\theta^0$, the derivatives of $g_{\theta, \rho}(Z_t, \theta, \gamma)$ satisfy integrability conditions sufficient for a uniform law of large numbers to apply and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{y}_{0t}(Z_t, \theta^0) \to^d N(0, \Sigma_{VR}^{LR})$. Then

$$\sqrt{T} (\hat{\theta} - \theta^0) \to N[0, S\Omega S]$$

where

$$\Omega = E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]' W_{22}^{-1} \Sigma_{22} W_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]$$

$$S = \left( E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]' W_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right] \right)^{-1}.$$ 

**Feasible efficient weighting matrices**

This section addresses estimation of an efficient weighting matrix. If $W_{22} = \Sigma_{22}$, then by Theorem 2.2.3

$$S\Omega S = \left( E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]' \Sigma_{22}^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right] \right)^{-1}$$
the efficiency bound of (1.21). So the problem is to estimate a matrix \( W \) with the property that:

\[
BW^{-1}B' = \begin{pmatrix}
W_{11} & W_{12} \\
W_{21} & \Sigma_{22}
\end{pmatrix}.
\]

Although this may seem daunting as the matrix \( B \) is not known to the researcher, characterizing such matrices is actually not difficult because of the special property of the function \( h^2(U, Z, \theta) \) from lemma 1.3.1 that \( h^2(U, Z, \theta^0) \) is \( Z \)-measurable. Recall that

\[
\Sigma_{22} = \text{Var}^{LR}_Q(h^2(U, Z, \theta^0)).
\]

Letting \( Q \) denote an arbitrary probability distribution on sequences of \((U_s, Z_s)\) such that the marginal distribution of \((Z_s)\) is equal to the true marginal distribution \((Z_s)\) given by the DGP. Using a subscript \( Q \) to denote variances and covariances with respect to this probability measure, the \( Z \)-measurability of \( h^2(U, Z, \theta^0) \) implies for any \((t, s)\)

\[
\text{Cov}_Q(h^2(U_t, Z_t, \theta^0), h^2(U_s, Z_s, \theta^0)) = \text{Cov}(h^2(U_t, Z_t, \theta^0), h^2(U_s, Z_s, \theta^0))
\]

and therefore \( \text{Var}_Q^{LR}(h^2(U_t, Z_t, \theta^0)) = \text{Var}^{LR}(h^2(U_t, Z_t, \theta^0)) = \Sigma_{22} \).

Thus

\[
W = \text{Var}_Q^{LR}(g(U, Z, \theta^0))^{-1} \quad \Rightarrow \quad BW^{-1}B' = B\text{Var}_Q^{LR}(g(U, Z, \theta^0))B' = \text{Var}_Q^{LR}(g^*(U, Z, \theta^0))
\]

The south-east block of \( \text{Var}_Q(g^*(U, Z, \theta^0)) \) is \( \text{Var}_Q(h^2(U, Z, \theta^0)) \). Therefore, this choice of \( W \) is an efficient weighting matrix. The question is then how to estimate the matrix \( \text{Var}_Q^{LR}(g(U, Z, \theta^0))^{-1} \).

This is not difficult as the researcher may choose to simulate values for the latent variables, \((\tilde{U}_s)\) from any convenient distribution and by construction the sequence of pairs obtained from observed \( Z \) and simulated \( \tilde{U} \), \((Z_s, \tilde{U}_s)\), has a joint distribution where the marginal distribution of \((Z_s)\) is equal to the true marginal distribution \((Z_s)\) given by the DGP. Then, \( \text{Var}_Q^{LR}(g(U, Z, \theta^0))^{-1} \) may be estimated using a preliminary consistent estimator of \( \theta^0 \), say obtained from an ELVIS estimator.
with $W = I$, combined with any heteroskedasticity and autocorrelation consistent (HAC) estimator.

### 2.3 Monte Carlo results

In this section I present some Monte Carlo results from the measurement error model discussed in chapter 1, subsection 1.2.1. Replications of size $T = 500$ are simulated from the data generating process:

$$x_t^* = \rho x_{t-1}^* + \delta_t,$$

$$\begin{pmatrix} 
\epsilon_t \\
u_t \\
\delta_t 
\end{pmatrix} \sim \text{IID } \mathcal{N} \left( \begin{pmatrix} 0 \\
0 \\
0 
\end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 - \rho^2 
\end{pmatrix} \right)$$

for various values of $\beta^0$ and with $\rho = 0.5$. For each replication I compute the two-step efficient ELVIS estimator using the moment restrictions (1.4) and the minimum distance estimator put forward by Gospodinov et al. (2017).

Gospodinov et al. (2017) suggest to estimate the measurement error model using a minimum distance approach. I simulate 1000 replications of the model for $\beta^0 = 1$, $\beta^0 = 1/4$, $\beta^0 = 1/8$, and $\beta^0 = 0$. The mean, standard deviation, and mean square error of the two-step efficient ELVIS estimator the Gospodinov et al. (2017) estimator are reported in table 2.1. Figure 2.1 contains a density plot of the results for $\beta^0 = 1$ and $\beta^0 = 1/4$, and figure 2.2 contains a density plot of the results for $\beta^0 = 1/8$ and $\beta^0 = 0$.

For replications where the parameter $\beta^0 = 1$, the ELVIS and minimum distance estimators perform similarly. However, the minimum distance estimator suffers from poor performance as the parameter approaches zero. Intuitively this is because the source of identification in the measurement error model is different when $\beta^0 = 0$ versus $\beta^0 \neq 0$. The Gospodinov et al. (2017) approach requires $\beta^0$ to be estimated jointly with other features of the data generating process. However, in the case when $\beta^0 = 0$, these other features are not identified and the estimator fails to have a Gaussian limiting distribution (see Phillips 1989). In finite samples the identification failure appears to impact estimation for values of $\beta^0$ near zero as well (see Andrews and Cheng 2012 and Han and McCloskey 2017). Figures 2.1 and 2.2 show that for smaller values of $\beta^0$ the sampling distribution of the
minimum distance estimator becomes extremely unstable with heavy tails and bimodalities.

2.4 Empirical example: long-run risk model

As an empirical exercise to illustrate the two-step efficient ELVIS estimator I consider the long-run risk model put forth by Bansal and Yaron (2004). This model posits that consumption and dividend growth rates contain a long-run predictable component. This predictable component is not observable by the researcher but does enter the decision problem facing economic agents. The model gives a potential resolution of equity premium puzzles by analyzing the risks that affect consumption.

As noted by Gospodinov et al. (2017), the constant volatility specification of the long-run risk model can be written in the framework of the classical measurement error model in chapter 1, subsection 1.2.1 as

\[ y_t = \mu_y + \beta x_t^* + u_t \]
\[ x_t = x_t^* + \epsilon_t \]

where \( y_t \) is the dividend growth rate, \( x_t \) is the consumption growth rate, and \( x_t^* \) is a latent predictable component of dividend and consumption growth. The parameter \( \beta \) is interpreted as the leverage ratio on expected consumption growth as in Abel (1999). Demeaning the data allows me to suppress the intercept and estimate the parameter \( \beta \) using the moment restrictions (1.4).

Data are provided publicly from Grammig and Küchlin (2018). I use quarterly U.S. data from the second quarter of 1947 to the fourth quarter of 2014. Consumption growth is measured from the Bureau of Economic Analysis as real personal consumption per capita of nondurable goods and services. Dividend growth is measured from the CRSP value-weighted market portfolio. Data are transformed into real terms using the consumer price index from the Bureau of Economic Analysis.

Estimation results are given in table 2.2. Column (1) presents estimation results for the two-step efficient ELVIS estimator using the moment restrictions (1.4). Columns (2) and (3) present the estimation results from the minimum distance estimator of Gospodinov et al. (2017) and results from the indirect inference estimator of Grammig and Küchlin (2018) respectively for comparison.

As expected the two-step efficient ELVIS estimator compares favorably with the extant methods in terms of standard errors. The point estimate obtained through the ELVIS estimator is consistent
with the results of the other methods as it falls within each method’s 95% confidence interval.

2.5 One-step efficient estimator for unconditional moment restrictions

In this section I examine one-step efficient estimators for unconditional moment models. In chapter 1 I have noted that the asymptotic efficiency bound for model (1.1) would be attained by the infeasible system of estimating equations:

\[
E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta^0) \right] Var(h_2(U_t, Z_t, \theta^0))^{-1} \frac{1}{T} \sum_{t=1}^{T} h_2(U_t, Z_t, \theta) = 0.
\]

As can be seen in the proof of Theorem 2.2.2 the asymptotic expansion for the two-step efficient ELVIS estimator takes this form. I now consider one-step estimators with asymptotic expansions that also take this form. Where the two-step approach of section 2.2 had clear parallels with efficient GMM, the estimators considered in this section are in the spirit of minimum empirical discrepancy estimators a la Corcoran (1998).

Newey and Smith (2004) has shown that empirical likelihood estimators exhibit higher order efficiency properties. One-step ELVIS estimators may inherit the desirable properties of one-step efficient methods without latent variables although I do not examine this question in detail here. Furthermore, Schennach (2014) has shown that set based inference may be conducted using one-step ELVIS objective functions of the form considered in this section for partially identified models. Thus, the result presented here is complementary as a researcher is assured that one-step ELVIS methods have appealing properties in either circumstance, point or set identification.

I consider minimum discrepancy type ELVIS estimators obtained through the optimization problem

\[
\min_{\pi_1, \ldots, \pi_T, \theta} \sum_{t=1}^{T} \varphi(\pi_t) \quad \text{Subject to} \quad \inf_{\gamma} \left\| \sum_{t=1}^{T} \pi_t \Psi_t, T, \Phi(\theta, \gamma)^{T} \right\| = 0 \quad (2.7)
\]

\[
\sum_{t=1}^{T} \pi_t = 1
\]
where

\[
\Psi_{t, \rho}(\theta, \gamma) = g_{\theta, \rho}(z_t, \theta, \gamma)
\]

\[
\Psi_{t, T, \rho}(\theta, \gamma) = \frac{1}{2S_n} \min(S_n, t - 1) \sum_{j=\max(-S_n, t-T)}^{t-1} \Psi_{t-j, \rho}(\theta, \gamma)
\]

and \(S_n \propto \sqrt{T}\). The optimization problem (2.7) has a strong resemblance to familiar estimation procedures put forth in, for example, Corcoran (1998) and Kitamura and Stutzer (1997). However the moment restrictions have been transformed by means of a preaveraging over a window determined by the bandwidth \(S_n\). The reason for this is to accommodate dependent data within a minimum discrepancy estimation framework. Here I focus on simple preaveraging to fix intuition but use of a kernel to obtain smoothed weights is also viable. Additional discussion and relevant technical conditions may be found in Kitamura (1997), Guggenberger and Smith (2008) and Smith (2011).

The function \(\varphi(\pi)\) is a differentiable convex function of a scalar; it is interpreted as a measure of the discrepancy between \(\pi\) and the empirical probability \(1/T\) of a single observation. I allow that \(\varphi(\pi)\) may depend on \(T\) but suppress this in the notation. When the solution to the optimization problem (2.7) \(\hat{\pi}_1, \ldots, \hat{\pi}_T, \hat{\theta}\) is unique with nonnegative \(\hat{\pi}_t\)s these may be interpreted as the probabilities that minimize discrepancy with the empirical probabilities subject to the moment restrictions. Particular choices of the function \(\varphi(\pi)\) result in generalized empirical likelihood estimators which include the well known cases of empirical likelihood, exponential tilting and Euclidean empirical likelihood or continuous updating efficient GMM.

The following result shows that the choice of dominating measure \(\rho\) is immaterial to the solution of the optimization problem (2.7): \(\hat{\pi}_1, \ldots, \hat{\pi}_T, \hat{\theta}\). It serves the same role in one-step estimation that Proposition 2.2.1 does for the two-step efficient ELVIS estimator.

**Proposition 2.5.1.** Suppose Assumption 2.2.1 holds. The set of pairs \((\theta, \pi)\) which satisfy the constraints in problem (2.7) is independent of the dominating measure \(\rho\) chosen by the researcher.

I therefore proceed as though \(\rho = P_{0,U|Z}\) (admissible by Assumption 2.2.2) so that the inf in the optimization problem (2.7) is attained asymptotically and so that the estimator’s asymptotic
behavior is described by the first order conditions of the Lagrangian:

\[ L = \sum_{t=1}^{T} \varphi(\pi_t) - \beta' T \sum_{t=1}^{T} \pi_t \Psi_{t,T,\rho}(\theta, \gamma) - \mu_T \left( \sum_{t=1}^{T} \pi_t - 1 \right). \] (2.8)

Before I can derive an expression for the first order conditions that characterize the estimator \( \hat{\theta} \), I define some preliminary notations: Note that for \( \pi_t \) nonnegative I have defined a probability distribution of the observable data. This combined with the family of conditional probability measures \( \nu(\cdot | Z, \theta, \gamma) \) given by

\[
\frac{d\nu(u|Z = z; \gamma, \theta)}{dP_0,U|Z}(u|Z = z) = \exp(\gamma' g(u, z, \theta)) \int \exp(\gamma' g(u, z, \theta)) d\mu_0^0(u|Z = z, \theta)
\] (2.9)

defines, for every \( (\theta, \gamma) \in \Theta \times \mathbb{R}^H \) joint probability distributions for the observed and latent variables.

I have, using obvious notations

\[
E_\pi[\text{Var}_\nu(g(U, Z, \theta)|Z)] = \sum_{t=1}^{T} \pi_t \int g(u, z_t, \theta)g(u, z_t, \theta)' \frac{\exp(\gamma' g(u, z_t, \theta))}{\int \exp(\gamma' g(u, z_t, \theta)) d\mu_0^0(u|z_t, \theta)} d\mu_0^0(u|Z = z_t, \theta)
\]

\[
\text{Var}_\pi^{LR}[E_\nu(g(U, Z, \theta)|Z)] = \sum_{t=1}^{T} \pi_t \Psi_{t,T,\rho}(\theta, \gamma) \Psi_{t,T,\rho}'(\theta, \gamma)
\]

where the dependence of \( \nu \) and \( \pi \) on \( \theta, \gamma \) has been made implicit to economize notation. A straightforward computation reveals that:

\[
E_\pi \left[ \frac{\partial \Psi_{t,T,\rho}(\theta^0, 0)}{\partial \gamma'} \right] = E_\pi[\text{Var}_\nu(g(U, Z, \theta)|Z)]. \] (2.10)

Using the same arguments as chapter 1, subsection 1.2.2, for each \( (\theta, \gamma) \) I may define an invertible linear transformation \( \tilde{N}_T(\theta, \gamma) \) such that

\[
E_\pi[\text{Var}_\nu(\tilde{N}_T(\theta, \gamma)g(U, Z, \theta)|Z)] = \begin{pmatrix} \hat{\Lambda} & 0 \\ 0 & 0 \end{pmatrix}
\] (2.11)

\[
\text{Var}_\pi^{LR}[E_\nu(\tilde{N}_T(\theta, \gamma)g(U, Z, \theta)|Z)] = \begin{pmatrix} \hat{\Sigma}_{11} & 0 \\ 0 & \hat{\Sigma}_{22} \end{pmatrix}
\] (2.12)

where \( \hat{\Lambda} \) is diagonal positive definite matrix.
The reference measure (2.9) implies that for consistent $(\hat{\theta}, \hat{\gamma})$:

$$E_\pi[\text{Var}_\nu(g(U, Z, \theta)| Z)] \to^p \Sigma_E$$

$$\text{Var}_\pi^L[E_\nu(g(U, Z, \theta)| Z)] \to^p \Sigma^L_{\nu}.$$ $$\text{Var}_\pi^R[E_\nu(g(U, Z, \theta)| Z)] \to^p \Sigma^R_{\nu}.$$ 

So by the maintained assumption that $\Sigma^L_{\nu}$ nonsingular, for large enough $T$ the matrix $\text{Var}_\pi^L(g(U, Z, \theta))$ is nonsingular and therefore $\hat{\Sigma}_{22}^{-1}$ exists with probability approaching one.

Making use of the matrix $\hat{N}_T$, I define the following notations

$$(\hat{N}_T(\hat{\theta}, \hat{\gamma})^{-1/2} \hat{\beta}_T) = (\hat{\beta}_{T, 1}, \hat{\beta}_{T, 2})$$

$$\hat{N}_T(\hat{\theta}, \hat{\gamma})\Psi_{t, T, \rho}(\hat{\theta}, \hat{\gamma}) = \begin{pmatrix} \Psi_{t, T, \rho, 1}(\hat{\theta}, \hat{\gamma}) \\ \Psi_{t, T, \rho, 2}(\hat{\theta}, \hat{\gamma}) \end{pmatrix}$$

where $\hat{\beta}_T$ is value of the Lagrange multiplier $\beta_T$ that minimizes (2.8).

**Theorem 2.5.1.** Suppose Assumptions 2.2.1 and 2.2.2 hold. Furthermore, assume that the infimum in (2.7) is attained and (2.7) uniquely defines estimators $(\hat{\theta}, \hat{\pi})$ with nonnegative $\hat{\pi}_t$s. Then $\hat{\theta}$ is characterized as the solution to the first order conditions:

$$\left( \sum_{t=1}^T \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial}{\partial \theta} \Psi_{t, T, \rho, 2}(\hat{\theta}, \hat{\gamma}) \right) \left( \sum_{t=1}^T \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \Psi_{t, T, \rho, 2}(\hat{\theta}, \hat{\gamma}) \Psi_{t, T, \rho, 2}(\hat{\theta}, \hat{\gamma})' \right)^{-1}$$

$$\sum_{t=1}^T \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \varphi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma})) \Psi_{t, T, \rho, 2}(\hat{\theta}, \hat{\gamma}) = 0$$

where $\varphi(\cdot)$ is the first derivative of the function $\varphi$.

The assumption that the infimum in (2.7) is attained is fairly high level but since I may assume without loss of generality that the $\rho = F_{0, U|Z}$ it is not restrictive in large samples.

**Lemma 2.5.1.** Assume that $\hat{\beta}_T = o_p(1)$, $\varphi(\cdot)$ is strictly convex, twice continuously differentiable, and $\varphi_{\pi}(p) \neq 0$ for $p > 0$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \varphi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma})) \Psi_{t, T, \rho}(\hat{\theta}, \hat{\gamma})$$
is proportional to
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) + o_p(1).
\]

Combining Lemma 2.5.1 with Theorem 2.5.1 I obtain that the GEL-ELVIS estimator for \( \hat{\theta} \) is characterized by:
\[
\left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial}{\partial \theta} \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \right) \left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) + o_p(1) = 0.
\]

To fix intuition I rewrite the previous equation as:
\[
E_{\pi} \left( \frac{\partial}{\partial \theta} \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \right) \left[ \text{Var}_{\pi} \left( \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma})' \right) \right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) + o_p(1) = 0.
\]

Hence the similarity of the minimum discrepancy first order conditions with the efficient estimating equations is already apparent. It remains to make the link between \( \Psi_{t,T,\rho,2} \) and \( h_2 \) explicit.

Note that \( \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) = B_2 \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \) where \( B_2 \) is an orthogonal matrix of eigenvectors for \( E_{\pi} [\text{Var}_\rho(g(U, Z, \hat{\theta}) | Z)] \) corresponding to the eigenvalue 0. This implies that \( \hat{B}_2 \hat{B}_2 \) is the projection onto the eigenspace on \( E_{\pi} [\text{Var}_\rho(g(U, Z, \hat{\theta}) | Z)] \) for the eigenvalue 0, and \( \hat{B}_2 \hat{B}_2 \) is an identity matrix. Similarly, \( h_2(U, Z, \theta) = B_2 g(U, Z, \theta) \) where \( B_2 \) is an orthogonal matrix of eigenvectors for \( \Sigma_E \) corresponding to the eigenvalue 0. Therefore \( B_2^T B_2 \) is the projection onto the eigenspace of \( \Sigma_E \) for the eigenvalue 0, and \( B_2 B_2^T \) is an identity matrix.

Then to make the link between \( B_2 \) and \( \hat{B}_2 \) I make use of the main result of Tyler (1981). Tyler uses the following assumption:

**Assumption 2.5.1.**

\[
\Pr [\text{rank}(E_{\pi} [\text{Var}_\rho(g(U, Z, \hat{\theta}))]) = \text{rank}[\Sigma_E]] \to 1
\]
\[
E_{\pi} [\text{Var}_\rho(g(U, Z, \theta) | Z)] \to \Sigma_E
\]

The main result of Tyler (1981) is that, under Assumption 2.5.1, \( \text{plim} \hat{B}_2 \hat{B}_2 = B_2^T B_2 \) (i.e. the projections onto the eigenspace converge). As noted above, the second condition will be satisfied
for consistent estimates of the parameters if I assume (without loss of generality) that the reference measure is given by \( P_{0,U|Z} \). The first condition is a requirement that the number of linearly independent \( Z \)-measurable linear combinations of \( g(U, Z, \hat{\theta}_T) \) is equal to the number of linearly independent \( Z \)-measurable linear combinations of \( g(U, Z, \theta^0) \) with probability approaching one. An interpretation of this condition is obtained from the obvious sufficient condition that for consistent \( \hat{\theta} \) Assumption 2.5.1 holds if

\[
\text{rank}(E_{\pi}[\text{Var}_{\nu}(g(U, Z, \theta))]) = c, \quad \forall \theta \in N_{\theta^0}
\]

where \( c \) is a constant and \( N_{\theta^0} \) is some neighborhood of \( \theta^0 \).

**Lemma 2.5.2.** Suppose Assumption 2.5.1 holds. The matrix \( D_T = B_2 \hat{B}_2' \) is invertible with probability approaching one.

Another immediate consequence of Tyler’s result is that:

\[
D_T \hat{B}_T = B_2 \hat{B}_2' \hat{B}_2 \rightarrow (B_2 B_2^2)^{-1} B_2 = B_2.
\]

Returning to the first order conditions, for sufficiently large \( T \) I may use the invertibility of \( D_T \) from Lemma 2.5.2 to write:

\[
\left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial}{\partial \theta} \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \right)^{\top} \left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma}) \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma})^2
\]

\[
= \left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial}{\partial \theta} (B_2 \Psi_{t,T,\rho})(\hat{\theta}, \hat{\gamma}) \right)^{\top} \left( \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) B_2 \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})(B_2 \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}))^2 \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} B_2 \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}).
\]
With this result in hand, an asymptotic expansion around \((\theta^0, 0)\) yields:

\[
E \left[ \frac{\partial h^r(U, Z, \theta^0)}{\partial \theta} \right] \left[ \frac{1}{T} \sum_{t=1}^{T} B_{2} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} B_{2} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})
\]

\[
\approx -E \left[ \frac{\partial h^r(U, Z, \theta^0)}{\partial \theta} \right] \left[ \frac{1}{T} \sum_{t=1}^{T} B_{2} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \right]^{-1} E \left[ \frac{\partial h_2(Z, U, \theta^0)}{\partial \theta} \right] \sqrt{T}(\hat{\theta} - \theta^0)
\]

\[
- E \left[ \frac{\partial h^r(U, Z, \theta^0)}{\partial \theta} \right] \left[ \frac{1}{T} \sum_{t=1}^{T} B_{2} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \right]^{-1} E \left[ \frac{\partial B_{2} \Psi_{t,T,\rho}(\theta^0, 0)}{\partial \gamma} \right] \sqrt{T}(\hat{\gamma} - \gamma^0).
\]

Finally, I deduce that the second term in the asymptotic expansion is asymptotically trivial as:

\[
E \left[ \frac{\partial B_{2} \Psi_{t,T,\rho}(\theta^0, 0)}{\partial \gamma} \right] = E[Cov(h_2(U, Z, \theta^0), g(U, Z, \theta^0)|Z)] = 0.
\]

### 2.6 One-step estimator for conditional moment restrictions

In this section I consider the problem of efficient estimation in model (1.2). Recall that in chapter 1 I have observed that the efficiency bound would be attained if the estimating equations

\[
\frac{1}{T} \sum_{t=1}^{T} E \left[ \frac{\partial h_{2,t-1}(U_t, Z_t, \theta^0)}{\partial \theta} \right] Var \left[ h_{2,t-1}(U_t, Z_t, \theta^0)|I_{t-1} \right]^{-1} h_{2,t-1}(U_t, Z_t, \theta) = 0
\]

could be implemented. Looking at these estimating equations, two clear challenges present themselves. The first is that the researcher does not have knowledge of the functions \(h_{2,t-1}\). As we have seen, properly implemented ELVIS type estimation procedures automatically extract these important directions from the moment function \(g\). Following this intuition, I propose a local generalized empirical likelihood (local GEL) type ELVIS estimator. Local GEL estimators have been used in econometrics to estimate conditional moment restriction models without latent variables using cross sectional data by Kitamura et al. (2004), Smith (2007), and Antoine et al. (2007). More recently, Gospodinov and Otsu (2012) have developed theory for local GEL using dependent data.

The second challenge with the estimating equations above is the presence of the conditional expectations. As is typical in the local GEL literature I will maintain the following assumption which allows conditional expectations to be estimated using kernel density methods.

**Assumption 2.6.1.** The data generating process of \((U_t, Z_t)\) is Markov of order 1.

Although in this exposition the DGP is Markov of order 1, higher order Markov models could be
handled analogously using kernel density methods. Using Assumption 2.6.1 I may rewrite (1.2) as:

\[ E[g(U_t, Z_t, \theta^0)|Z_{t-1}] = 0. \]  

(2.13)

Let

\[ w_{tj} = \frac{K((X_{t-1} - X_j)/b_T)}{\sum_{i=1}^{T} K((X_{t-1} - X_i)/b_T)} \]

denote kernel weights for local smoothing, where \( K \) is a kernel function and \( b_T \) is a bandwidth parameter. Additional discussion and relevant technical conditions is found in Gospodinov and Otsu (2012).

The local GEL ELVIS estimator is defined by the optimization problem:

\[
\min_{\pi_{11}, \ldots, \pi_{TT}, \theta} \sum_{t=1}^{T} \sum_{j=1}^{T} w_{tj} \varphi(\pi_{tj}/w_{tj}) \quad \text{Subject to} \\
\inf_{\gamma_t} \left\| \sum_{j=1}^{T} \pi_{tj} \Psi_j(\theta, \gamma_t) \right\| = 0 \quad \forall t \in \{1, \ldots, T\} \\
\sum_{j=1}^{T} \pi_{tj} = 1 \quad \forall t \in \{1, \ldots, T\} \tag{2.14}
\]

where

\[ \Psi_j(\theta, \gamma_t) = g_{\theta, \rho_t}(z_j, \theta, \gamma_t) \]

\[ \varphi(\pi/w) = \begin{cases} 
\frac{1}{1-\lambda} \frac{\pi^{1-\lambda}}{w^{1-\lambda}}, & \lambda \neq 0, 1 \\
-\log(\pi/w), & \lambda = 1 
\end{cases} \]

The family of functions \( \varphi(\cdot) \) considered here belong to Cressie-Read family of divergences. These divergences are a subset of the functions that were considered in the unconditional case. The Cressie-Read divergences considered here will facilitate interpretation of the resulting estimating equations through the property that:

\[ \frac{\partial \varphi(\pi/w)}{\partial \pi} = \varphi(\pi/w) = \frac{\pi^{-\lambda}}{w^{1-\lambda}}. \]
Similar to the unconditional case in section 2.5, \( \varphi(\cdot) \) is a differentiable convex function of a scalar and is interpreted as a measure of the discrepancy between \( \pi_{tj} \) and the smoothed empirical probability \( w_{tj} \). When the solution to the optimization problem (2.14) \( \hat{\pi}_{11}, \ldots, \hat{\pi}_{TT}, \hat{\theta} \) is unique with nonnegative \( \hat{\pi}_{tj} \)s these may be interpreted as the conditional probabilities that minimize the discrepancy with the smoothed empirical probabilities subject to the conditional moment restrictions. In light of this interpretation, for any integrable function \( q(Z) \) the quantity

\[
\sum_{j=1}^{T} \pi_{tj} q(Z_j)
\]

defines an estimator of the conditional expectation \( E[q(Z_t)|Z_{t-1} = z_{t-1}] \) that incorporates information contained within the conditional moment restrictions. I denote this constrained estimator as:

\[
\hat{E}_T[q(Z_t)|Z_{t-1} = z_{t-1}] = \sum_{j=1}^{T} \pi_{tj} q(Z_j)
\]

Contrast this with the unconstrained Nadaraya-Watson estimator of the same conditional expectation which is given by \( \sum_{j=1}^{T} w_{tj} q(Z_j) \).

In addition to the local smoothing, the conditional moment restrictions require that a different set of nuisance parameters be allowed for every observed value of the conditioning variable. Hence the subscript in \( \gamma_t \) above. Furthermore, I allow for the possibility that a different dominating measure is used for each value of the conditioning variable and therefore denote the dominating measure with a subscript as \( \rho_t \). I will denote the true probability distribution of the latent variable \( U_t \) given the observed values of \( Z_t \) and \( Z_{t-1} \) as \( \mu_t^0 \).

The following assumptions about the dominating measures should appear familiar from the discussion of estimation in model (1.1). These assumptions serve the same purpose as their counterparts from the unconditional model.

**Assumption 2.6.2.** The user specified reference measures have support that contains the true support of \( U_t|Z_t, Z_{t-1} \) and are such that

\[
\int \log \left( \int \exp(\gamma' g(u, z, \theta)) \rho_t(du|z; \theta) \right) P_Z(dz)
\]
exists and is twice continuously differentiable in $\gamma$ for all $\theta \in \Theta$ and $\gamma \in \mathbb{R}^H$.

The following proposition summarizes the important fact that, under Assumption 2.6.2, the solution $\hat{\theta}$ of the constrained optimization problem (2.14) is numerically independent of the particular reference measures $\rho_t(\cdot|z;\theta)$ that are used.

**Proposition 2.6.1.** Suppose Assumption 2.6.2 holds for $\rho_t$ for $1 \leq t \leq T$. The set of pairs $(\theta, \pi)$ which satisfy the constrains in problem (2.14) is independent of the dominating measures $\rho_t$ chosen by the researcher.

**Assumption 2.6.3.** The true conditional probabilities of $U_t$ given $Z_t, Z_{t-1}, \mu^0_t$, are such that

$$\int \log \left( \int \exp(\gamma' g(u, z, \theta)) \mu^0_t(du|z) \right) P_Z(dz) \quad (2.15)$$

exists and is twice continuously differentiable in $\gamma$ for all $\theta \in \Theta$ and $\gamma \in \mathbb{R}^H$.

Assumption 2.6.3 says that the true conditional distributions satisfy the conditions of Assumption 2.6.2. In view of Proposition 2.6.1 and Assumption 2.6.3 I may proceed as if $\rho_t = \mu^0_t$ so that the inf in the optimization problem is attained asymptotically and the estimators asymptotic behavior is described by the first order conditions of the Lagrangian:

$$L = \sum_{t=1}^{T} \left( \sum_{j=1}^{T} w_{tj} \varphi(\pi_{tj}/w_{tj}) - \beta_t' \sum_{j=1}^{T} \pi_{tj} \Psi_j(\theta, \gamma_t) - \mu_t \left( \sum_{j=1}^{T} \pi_{tj} - 1 \right) \right).$$

I begin by deriving the first order conditions for this estimator of $\theta^0$. Note that for $\pi_{tj}$'s non-negative a solution to problem (2.14) defines a conditional probability distribution of the observable data. This combined with the families of conditional probability measures $\nu_t(\cdot|z, \theta, \gamma_t)$ given by

$$\frac{d\nu_t(u|z; \gamma_t, \theta)}{d\mu^0_t(u|z)} = \frac{\exp(\gamma'_t g(u, z, \theta))}{\int \exp(\gamma'_t g(u, z, \theta)) d\mu^0_t(u|z)} \quad (2.16)$$

defines, for every $(\theta, \gamma_t) \in \Theta \times \mathbb{R}^H$ a joint probability distribution for the observed data and the latent variables. Using obvious notations, a variance decomposition is given by

$$Var_{\pi \times \nu_t}(g(U_t, Z_t, \theta)|Z_{t-1} = z_{t-1}) = E_{\pi}[Var_{\nu_t}(g(U_t, Z_t, \theta)|Z_t, Z_{t-1})|Z_{t-1} = z_{t-1}]$$

$$+ Var_{\pi}[E_{\nu_t}(g(U_t, Z_t, \theta)|Z_t, Z_{t-1})|Z_{t-1} = z_{t-1}]$$
where the dependence of \( \nu_t \) and \( \pi \) on \( \theta, \gamma_t \) has been made implicit to economize notation. For each \((\theta, \gamma_t)\) I can now define a \( Z_{t-1} \)-measurable invertible linear transformation \( \hat{N}_T^{-1}(\theta, \gamma_t) \) such that:

\[
E_{\pi} [ \text{Var}_{\nu_t} (\hat{N}_T^{-1}(\theta, \gamma_t) g(U_t, Z_t, \theta)|Z_t, Z_{t-1}) | Z_{t-1} = z_{t-1}] = \begin{pmatrix} \hat{\Lambda}^{-1} & 0 \\ 0 & 0 \end{pmatrix} (2.17)
\]

\[
\text{Var}_{\pi} [E_{\nu_t} (\hat{N}_T^{-1}(\theta, \gamma_t) g(U_t, Z_t, \theta)|Z_t, Z_{t-1}) | Z_{t-1} = z_{t-1}] = \begin{pmatrix} \hat{F}_1^{-1} & 0 \\ 0 & \hat{\Sigma}_2^{-1} \end{pmatrix}. (2.18)
\]

Where \( \hat{\Lambda}^{-1} \) is diagonal positive definite diagonal matrix and \( \hat{\Sigma}_2^{-1} \) is positive definite with probability approaching one. Making use of the matrix \( \hat{N}_T^{-1} \), I define the following notations:

\[
(\hat{N}_T^{-1}(\hat{\theta}, \hat{\gamma})^{-1})' = (\hat{\beta}_{\gamma 1}, \hat{\beta}_{\gamma 2})
\]

\[
\hat{N}_T^{-1}(\hat{\theta}, \hat{\gamma}) \Psi_{\gamma}(\hat{\theta}, \hat{\gamma}) = \begin{pmatrix} \Psi_{\gamma,1}(\hat{\theta}, \hat{\gamma}) \\ \Psi_{\gamma,2}(\hat{\theta}, \hat{\gamma}) \end{pmatrix}.
\]

With these notations in place the first order conditions of the Lagrangian lead to the following

**Theorem 2.6.1.** Suppose Assumptions 2.6.2 and 2.6.3 hold. Furthermore, assume that the infimum in (2.14) is attained and (2.14) uniquely defines estimators \((\hat{\theta}, \hat{\gamma})\) with nonnegative \( \hat{\pi}_{ij} \)'s. Then \( \hat{\theta} \) is characterized as the solution to the first order conditions:

\[
\sum_{t=1}^{T} \left( \sum_{j=1}^{T} \frac{\partial}{\partial \theta} \Psi_{j,2}(\hat{\theta}, \hat{\gamma}) \right) \left( \sum_{j=1}^{T} \frac{\partial}{\partial \theta} \Psi_{j,2}(\hat{\theta}, \hat{\gamma}) \Psi_{j,2}(\hat{\theta}, \hat{\gamma})' \right)^{-1} \left( \sum_{j=1}^{T} \hat{\nu}_{tj} \Psi_{j,2}(\hat{\theta}, \hat{\gamma})' \right) = 0.
\]

To facilitate interpretation, the expression of Theorem 2.6.1 can be rewritten as:

\[
\sum_{t=1}^{T} \hat{E}_T \left[ \frac{\partial}{\partial \theta} \Psi_{j,2}(\hat{\theta}, \hat{\gamma}) | Z_{t-1} = z_{t-1} \right] \hat{E}_T \left[ \Psi_{j,2}(\hat{\theta}, \hat{\gamma}) \Psi_{j,2}(\hat{\theta}, \hat{\gamma})' | Z_{t-1} = z_{t-1} \right]^{-1} \left( \sum_{j=1}^{T} \hat{\nu}_{tj} \Psi_{j,2}(\hat{\theta}, \hat{\gamma})' \right) = 0.
\]

Since both \( \hat{\nu}_{ij} \) and \( w_{ij} \) are both localization weights that can be used to estimate conditional expectations given \( Z_{t-1} = z_{t-1} \), the last factor in the previous equation can also be interpreted as
an estimator for a conditional moment. I will denote this:

\[ \hat{E}_T \left[ \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t) | Z_{t-1} = z_{t-1} \right] = \sum_{j=1}^{T} w_j^A \pi_j (\hat{\theta}, \hat{\gamma}_t)^{1-\lambda} \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t). \]

This estimator is equal to the Nadaraya-Watson estimator when \( \lambda = 1 \). With these notations in place, the first order conditions can be interpreted as:

\[
\frac{1}{T} \sum_{t=1}^{T} \hat{E}_T \left[ \frac{\partial}{\partial \theta} \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t) | Z_{t-1} = z_{t-1} \right] \hat{E}_T \left[ \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t) \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t)' | Z_{t-1} = z_{t-1} \right]^{-1} - \hat{E}_T \left[ \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t) | Z_{t-1} = z_{t-1} \right] = 0.
\]

Next is to make the link between \( \Psi_{t,2} \) and \( h_{2,t-1} \). The argument is the same as in Section 2.5 up to a change in notations. Once again \( \Psi_{t,2}(\hat{\theta}, \hat{\gamma}_t) = \hat{B}_2^{-1} \Psi_t(\hat{\theta}, \hat{\gamma}_t) \) where \( \hat{B}_2^{-1} \) is an orthogonal matrix of eigenvectors for

\[ \hat{\Sigma}_{E}^{-1} = \hat{E}_t [Var_{\nu_t}(g(U_t, Z_t, \hat{\theta}) | Z_t, Z_{t-1} = z_{t-1})] \]

corresponding to the eigenvalue 0. The produce \( \hat{B}_2^{-1} \hat{B}_2^{-1} \) is the projection onto the eigenspace of \( \hat{\Sigma}_{E}^{-1} \) for the eigenvalue 0, and \( \hat{B}_2^{-1} \hat{B}_2^{-1} \) is an identity matrix. Similarly, \( h_{2,t-1}(U, Z, \theta) = B_2^{-1} g(U_t, Z_t, \theta) \) where \( B_2^{-1} \) is an orthogonal matrix of eigenvectors for \( \Sigma_{E}^{-1} \) corresponding to the eigenvalue 0. So that \( B_2^{-1} \) is the projection onto the eigenspace on \( \Sigma_{E}^{-1} \) for the eigenvalue 0, and \( B_2^{-1} B_2^{-1} \) is an identity matrix. The assumption needed to use the result of Tyler (1981) is then:

**Assumption 2.6.4.** For all \( z_{t-1} \) is the support of \( Z \):

\[
P[rank(\hat{E}_t [Var_{\nu_t}(g(U, Z, \hat{\theta}) | Z_t, Z_{t-1} = z_{t-1})]) | Z_{t-1} = z_{t-1})] = rank(\Sigma_{E}^{-1}) \to 1
\]

\[
\hat{E}_t [Var_{\nu_t}(g(U, Z, \theta) | Z_t, Z_{t-1} = z_{t-1})] | Z_{t-1} = z_{t-1}) \to \Sigma_{E}^{-1}.
\]

Under Assumption 2.6.4 \( \hat{B}_2^{-1} \hat{B}_2^{-1} \to B_2^{-1} \hat{B}_2^{-1} \). The interpretation of Assumption 2.6.4 is analogous to Assumption 2.5.1 is section 2.5. The following lemma is reproduced from the unconditional case.

**Lemma 2.6.1.** Suppose Assumption 2.6.4 holds. The matrix \( D_t^{-1} = B_2^{-1} \hat{B}_2^{-1} \) is invertible with
probability approaching one.

Note also that

\[ D_T^{-1} B_T^{-1} = B_T^{-1} B_T^{-1} \rightarrow (B_T^{-1} B_T^{-1}) B_T^{-1} = B_T^{-1}. \]

Returning to the first order conditions, for sufficiently large \( T \), the invertiblity of \( D_T^{-1} \) implies:

\[
\begin{align*}
&\sum_{t=1}^{T} \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial}{\partial \theta} \Psi_j(\hat{\theta}, \hat{\gamma}_t) \right) \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)' \right)^{-1} \left( \sum_{j=1}^{T} w_{tj} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t) \right) \\
&= \sum_{t=1}^{T} \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial}{\partial \theta} \Psi_j(\hat{\theta}, \hat{\gamma}_t) \right) (D_T^{-1} B_T^{-1})' \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)' (D_T^{-1} B_T^{-1})' \right)^{-1} \\
&= \sum_{t=1}^{T} \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial}{\partial \theta} (B_T^{-1} \Psi_j)'(\hat{\theta}, \hat{\gamma}_t) \right) \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) B_T^{-1} \Psi_j(\hat{\theta}, \hat{\gamma}_t)' B_T^{-1} \right)^{-1} \\
&= \sum_{j=1}^{T} w_{tj} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)' + o_p(1).
\end{align*}
\]

With the first order conditions in this form an asymptotic expansion around \((\theta^0, 0)\) yields:

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right] \left[ \Sigma_{22}^{t-1} \right]^{-1} h_{2,t-1}(U_t, Z_t, \theta^0) \\
\approx -E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right]' \left[ \Sigma_{22}^{t-1} \right]^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right] \sqrt{T} (\hat{\theta} - \theta^0) \\
- \sum_{t=1}^{T} E \left[ \frac{\partial}{\partial \gamma_t} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right]' \left[ \Sigma_{22}^{t-1} \right]^{-1} E \left[ \frac{\partial}{\partial \gamma_t} B_T h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right] \sqrt{T} \hat{\gamma}.
\]
Finally, noting that

\[ E \left[ \frac{\partial}{\partial \gamma_t} B_t^{t-1} \Psi_t(\theta^0, \gamma_t^0) | Z_{t-1} \right] = E \left[ \text{Cov}(h_{2,t-1}(U_t, Z_t, \theta^0), g(U_t, Z_t, \theta^0)) | Z_t, Z_{t-1} \right] = 0 \]

the asymptotic expansion becomes

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right] \left[ \Sigma_{22}^{-1} \right]^{-1} h_{2,t-1}(U_t, Z_t, \theta^0) \\
\approx -E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right]^{\prime} \left[ \Sigma_{22}^{-1} \right]^{-1} E \left[ \frac{\partial}{\partial \theta} h_{2,t-1}(U_t, Z_t, \theta^0) | Z_{t-1} \right] \sqrt{T}(\hat{\theta} - \theta^0)
\]

so that the expected asymptotic distribution follows.

### 2.7 Conclusion

In this chapter I have presented a framework for efficient semiparametric estimation of models that contain latent variables. The models are defined in terms of some moment restrictions either conditional or unconditional. The estimation procedures I proposed are inspired by the entropic latent variable integration via simulation framework of Schennach (2014). One-step estimators in the spirit of generalized empirical likelihood and two-step estimators in spirit of generalized method of moments were considered.

In the important example of measurement error both Monte Carlo simulations and an empirical example provide evidence that the estimators discussed perform well in practice. Other avenues for future empirical research include dynamic factor analysis models, models of correlation between markets, and multivariate volatility models.
2.8 Appendix: Proofs

2.8.1 Proofs of Propositions

Proof of Proposition 2.2.1:

Proof. Corollary 2.1 from Schennach (2014) says that under Assumption 2.2.1, for any \( \theta \in \Theta \) and for any \( R \in \mathcal{P}_Z \),

\[
\left\{ \int \int g(u, z, \theta)Q(du|z) R(dz) | Q \in \mathcal{P}_{U|Z} \right\} = \left\{ \int g_{\theta, \rho}(z, \theta, \gamma) R(dz) | \gamma \in \mathbb{R}^H \right\}
\]

where \( \mathcal{P}_{U|Z} \) denotes the set of all regular conditional probability measures with support equal to the support of \( U \) or any measurable subset of the support of \( U \) and \( \mathcal{P}_Z \) denotes the set of all probability measures with support equal to the support of \( Z \) or any measurable subset of the support of \( Z \). Note that the sets in the previous display are subsets of \( \mathbb{R}^H \) and that if the measure \( R \) is such that some component of \( g(u, z, \theta) \) is not integrable for any \( Q \in \mathcal{P}_{U|Z} \) than the sets are empty.

Possible choices for \( R \) in this result include the empirical distribution of \( Z \) for a given sample: in this case

\[
\left\{ \frac{1}{T} \sum_{t=1}^{T} g(u, Z_t, \theta)Q(du|z) | Q \in \mathcal{P}_{U|Z} \right\} = \left\{ \frac{1}{T} \sum_{t=1}^{T} \tilde{g}_{\rho}(Z_t, \theta, \gamma) | \gamma \in \mathbb{R}^H \right\}.
\]

(2.19)

Since the set on the left hand side is independent of \( \rho \), the set on the right hand side does not depend on \( \rho \) either. In particular, the value of the profiled objective function is

\[
\inf_{\gamma \in \mathbb{R}^p} \hat{Q}_T(\theta, \gamma) = \inf_{\gamma \in \mathbb{R}^p} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{g}_{\rho}(Z_t, \theta, \gamma) \right)' \hat{W} \left( \frac{1}{T} \sum_{t=1}^{T} \tilde{g}_{\rho}(Z_t, \theta, \gamma) \right)
\]

where \( \| \cdot \|_{\hat{W}} \) denotes the semi-norm induced by the positive semi-definite matrix \( \hat{W} \). That is, the profiled objective function for any \( \theta \in \Theta \) is exactly equal to the smallest norm of all vectors in the set of display (2.19) and is therefore independent of the reference measure. Finally, the extremum
estimator of $\theta^0$ is

$$\hat{\theta} = \arg \min_{\theta} \inf_{\gamma \in \mathbb{R}^p} \hat{Q}(\theta, \gamma)$$

from which the announced result follows. \qed

**Proof of Proposition 2.5.1**

*Proof.* In view of Schennach (2014) Theorem 2.1 the solution to the constrained maximization problem $(\hat{\theta}, \hat{\pi})$ is independent of the measure $\rho$ used. Indeed, up to change of notation, theorem 2.1 in Schennach (2014) states that if for some fixed $\rho = \bar{\rho}$ the vector $(\theta, \pi)$ is such that

$$\inf_{\gamma} \left\| \sum_{t=1}^{T} \pi_t \Psi_{t,T,\bar{\rho}}(\theta, \gamma) \right\| = 0$$

then it is also true for any other admissible $\tilde{\rho}$ that

$$\inf_{\gamma} \left\| \sum_{t=1}^{T} \pi_t \Psi_{t,T,\tilde{\rho}}(\theta, \gamma) \right\| = 0$$

thus the constraint set, and therefore the solution, of the minimum empirical discrepancy problem is independent of $\rho$. \qed

**Proof of Proposition 2.6.1**

*Proof.* For any $t$, Proposition 2.5.1 tells use that the constraints on $(\pi_{1t}, \ldots, \pi_{Tt}, \theta)$ imposed by

$$\sum_{j=1}^{T} \pi_{1j} = 1$$

$$\sum_{j=1}^{T} \pi_{tj} \Psi_j(\theta, \gamma_t) = 0$$

are independent of $\rho_t$. The proposition follows by noting that the constraint set in problem (2.14) is given by combining these constraints for all $t$. \qed
2.8.2 Proofs of Lemmata

Proof of Lemma 2.2.1

Proof. For any two vectors \( \gamma_1, \gamma_2 \) that satisfy equation (2.5) the Kullback-Leibler divergence of the associated probability measures is

\[
D(P_{\gamma_1}, P_{\gamma_2})
= \int \int (\gamma_1 - \gamma_2) g(u, z, \theta^0) - a(z, \theta^0, \gamma_1) + a(z, \theta^0, \gamma_1)) \rho(du|z; \gamma_1, \theta^0) P^z(dz)
= (\gamma_1 - \gamma_2) \int \int g(u, z, \theta^0) \rho(du|z; \gamma_1, \theta^0) P^z(dz)
+ \int \int (a(z, \theta^0, \gamma_2) - a(z, \theta^0, \gamma_1)) \rho(du|z; \gamma_1, \theta^0) P^z(dz)
= \int (a(z, \theta^0, \gamma_2) - a(z, \theta^0, \gamma_1)) P^z(dz)
\]

where the final equality follows because \( \gamma_1 \) satisfies equation (2.5). However a symmetric calculation shows that

\[
D(P_{\gamma_2}, P_{\gamma_1})
= \int (a(z, \theta^0, \gamma_1) - a(z, \theta^0, \gamma_2)) P^z(dz)
= -D(P_{\gamma_1}, P_{\gamma_2})
\]

so that by the non-negativity of the divergence, \( D(P_{\gamma_1}, P_{\gamma_2}) = 0 \).

Proof of Lemma 2.2.2:

Proof. Recall from lemma 2.2.1 that for any two vectors \( \gamma_1, \gamma_2 \in \mathcal{G}^0 \) the Kullback-Leibler divergence between the associated distributions is zero so that the associated densities are equal

\[
\frac{\exp(\gamma_1'g(u, z, \theta^0))}{\int \exp(\gamma_1'g(u, z, \theta^0)) \rho(du|z; \theta^0)} = \frac{\exp(\gamma_2'g(u, z, \theta^0))}{\int \exp(\gamma_2'g(u, z, \theta^0)) \rho(du|z; \theta^0)}
\]

almost surely. Rearranging the previous display:

\[
\log \left( \frac{\int \exp(\gamma_1'g(u, z, \theta^0)) \rho(du|z; \theta^0)}{\int \exp(\gamma_2'g(u, z, \theta^0)) \rho(du|z; \theta^0)} \right) = (\gamma_1 - \gamma_2)'g(u, z, \theta^0).
\]

(2.20)
Since the left side the of previous equation does not depend on $u$, the right hand side must not either. To show that $\mathcal{G}^0$ is convex it is sufficient to show that $\gamma_\alpha = \alpha \gamma_1 + (1 - \alpha) \gamma_2$ satisfies equation (2.5) for all $\alpha \in [0, 1]$. I compute:

$$\frac{\exp(\gamma'_o g(u, z, \theta^0))}{\int \exp(\gamma'_o g(u, z, \theta^0)) \rho(du|z; \theta^0)} = \frac{\exp((\alpha(\gamma_1 - \gamma_2) + \gamma_2')^T g(u, z, \theta^0))}{\int \exp((\alpha(\gamma_1 - \gamma_2) + \gamma_2')^T g(u, z, \theta^0)) \rho(du|z; \theta^0)}$$

$$= \frac{\exp(\alpha(\gamma_1 - \gamma_2') g(u, z, \theta^0)) \exp(\gamma_2' g(u, z, \theta^0))}{\int \exp(\alpha(\gamma_1 - \gamma_2') g(u, z, \theta^0)) \exp(\gamma_2' g(u, z, \theta^0)) \rho(du|z; \theta^0)}$$

$$= \frac{\exp(\alpha(\gamma_1 - \gamma_2') g(u, z, \theta^0)) \int \exp(\gamma_2' g(u, z, \theta^0)) \rho(du|z; \theta^0)}{\int \exp(\gamma_2' g(u, z, \theta^0)) \rho(du|z; \theta^0)}$$

where the third equality follows because $(\gamma_1 - \gamma_2') g(u, z, \theta^0)$ does not depend on $u$ from equation (2.20). Thus,

$$\int \int W g(u, z, \theta^0) \frac{\exp(\gamma'_o g(u, z, \theta^0))}{\int \exp(\gamma'_o g(u, z, \theta^0)) \rho(du|z; \theta^0)} \rho(du|z; \theta^0) P_Z^\mu (dz)$$

$$= \int \int W g(u, z, \theta^0) \frac{\exp(\gamma'_2 g(u, z, \theta^0))}{\int \exp(\gamma'_2 g(u, z, \theta^0)) \rho(du|z; \theta^0)} \rho(du|z; \theta^0) P_Z^\mu (dz)$$

$$= 0$$

because $\gamma_2 \in \mathcal{G}^0$. Therefore, $\gamma_\alpha \in \mathcal{G}^0$. \qed

**Proof of Lemma 2.5.1:**

Proof. The implied probabilities can be written as $\hat{\pi}_t = h^{-1}_\pi(\mu + \hat{\beta}_T \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma}))$ so that

$$\sum_{i=1}^T h^{-1}_\pi(\mu + \hat{\beta}_T \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma})) \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma}) = 0$$

and by a Taylor expansion around $\beta = 0$

$$h^{-1}_\pi(\mu) \sum_{i=1}^T \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma}) + \frac{1}{\varphi \pi(h^{-1}_\pi(\mu))} \sum_{i=1}^T \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,o}(\hat{\theta}, \hat{\gamma}) \hat{\beta}_T = o_p(1)$$
so that

$$
\sqrt{T} \hat{\beta}_T \propto \left( \frac{1}{T} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) + o_p(1)
$$

and by another Taylor expansion around \( \beta = 0 \)

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \varphi_\pi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma})) \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_\pi^{-1}(\mu + \hat{\beta}_T^\prime \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) (\mu + \hat{\beta}_T^\prime \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})) \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) + o_p(1)
$$

$$
= \frac{h_\pi^{-1}(\mu) \mu}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) + \left( \frac{1}{\varphi_\pi(h_\pi^{-1}(\mu))} + \mu \right) \frac{1}{T} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})' \sqrt{T} \hat{\beta}_T + o_p(1)
$$

$$
\propto \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma}) + o_p(1)
$$

\( \square \)

**Proof of Lemma 2.5.2:**

*Proof.* Note that

$$
D_T D_T' = B_2[B_2' B_2] B_2'
$$

And by main result of Tyler (1981):

$$
\text{plim} B_2' B_2 = B_2' B_2
$$

Therefore,

$$
\text{plim} B_2[B_2' B_2] B_2' \rightarrow (B_2 B_2') (B_2 B_2') = I^2 = I
$$

which implies \( D_T D_T' \) converges to matrix with full rank and therefore, \( D_T \) has full rank with probability approaching one.

\( \square \)
Proof of lemma 2.6.1

Proof. The proof is the same as that of lemma 2.5.2 up to an obvious change in notation.

2.8.3 Proofs of Theorems

Proof of Theorem 2.2.1:

The proof of Theorem 2.2.1 is obtained using the following lemma:

Lemma 2.8.1. Under Assumption 2.2.4, for any positive integer $k$ and all $\epsilon > 0$

$$P\left(\|\hat{\theta}_T\|^2 + \|\hat{\gamma}_T\|^2 \leq k^2 \text{ and } d((\hat{\theta}_T, \hat{\gamma}_T), \Delta_{0,k}) \geq \epsilon\right) \to 0$$

Proof. For any $\epsilon > 0, k \in \mathbb{N}$ I have that

$$P(d((\theta, \gamma), \Delta_{0,k}) \geq \epsilon) = P\left(\|\hat{\theta}_T\|^2 + \|\hat{\gamma}_T\|^2 \leq k^2 \text{ and } d((\hat{\theta}_T, \hat{\gamma}_T), \Delta_{0,k}) \geq \epsilon\right) + P\left(\|\hat{\theta}_T\|^2 + \|\hat{\gamma}_T\|^2 > k^2 \text{ and } d((\hat{\theta}_T, \hat{\gamma}_T), \Delta_{0,k}) \geq \epsilon\right) = (A) + (B)$$

Term (A) above converges to zero as $n \to \infty$ by Lemma 2.8.1. For term (B) note that

$$(B) = P(d((\theta, \gamma), \Delta_{0,k}) \geq \epsilon | \|\hat{\theta}_T\|^2 + \|\hat{\gamma}_T\|^2 > k^2) P(\|\hat{\theta}_T\|^2 + \|\hat{\gamma}_T\|^2 > k^2)$$

which converges to zero as $n \to \infty$ by the assumption that $P(\|\hat{\theta}\|^2 + \|\hat{\gamma}\|^2 > k^2) \to 0$.

To conclude that $\hat{\theta}_T \to^p \theta^0$ and $d(G^0, \hat{\gamma}_T) \to^p 0$ I will show that with probability approaching one $(\hat{\theta}_T, \hat{\gamma}_T)$ belongs to the open set:

$$U_\delta = \{(\theta, \gamma) \in \Delta : d((\theta, \gamma), (\theta^0, G^0)) < \delta\}$$

for all $\delta > 0$. Note that the set

$$C = U_{\delta/2}^c \cap (\Theta \times G) \cap \{(\theta, \gamma) : \|\theta_T\|^2 + \|\gamma_T\|^2 \leq k^2\}$$
is compact so that
\[
\sup_{(\theta, \gamma) \in C} Q_0(\theta, \gamma) = Q_0(\theta^*, \gamma^*) < \sup_{(\theta, \gamma) \in \Delta} Q_0(\theta, \gamma)
\]

For \( k \) sufficiently large that
\[
\sup_{(\theta, \gamma) \in \Delta} Q_0(\theta, \gamma) - Q_0(\theta^*, \gamma^*) > \frac{1}{k} \Rightarrow \Delta_{0,k} \subseteq U_{\delta/2}
\]

Now setting \( \epsilon = \frac{\delta}{2} \), with probability approaching one \((\hat{\theta}_T, \hat{\gamma}_T) \in U_{\delta}\).

**Proof of Theorem 2.2.2:**

The proof of Theorem 2.2.2 uses the lemma below.

**Lemma 2.8.2.** For any linear transformation \( S \) of the form

\[
S = \begin{pmatrix}
    Id & P \\
    0 & Id
\end{pmatrix} B
\]

where \( P \) is a \( r_E \times H - r_E \) matrix, there exists a family of conditional probability distributions \( \mu^S(\cdot|Z, \theta) \) such that \( \mu^S(\cdot|Z, \theta^0) = \mu^0(\cdot|Z) \) and

\[
SE\left( \frac{\partial g_{\theta, \mu^S}(Z, \theta^0, \gamma^0)}{\partial \theta} \right) = \left[ 0, E\left( \frac{\partial g_2(Z, \theta^0)'}{\partial \theta} \right) \right]'
\]

**Proof.** Let

\[
S = \begin{pmatrix}
    Id & -\Sigma_{12}\Sigma_{22}^{-1} \\
    0 & Id
\end{pmatrix} B
\]

and without loss of generality assume that the reference measure is given by the measure \( \mu^S \) from lemma 2.8.2 by Assumption 2.2.2. In light of this a psuedo-true value of \( \gamma \) is 0 and the following
identifies hold:

\[ g_{\theta, \rho}(z, \theta, 0) = \tilde{g}_0(z, \theta) \]

\[
E \left[ \frac{\partial g_{\theta, \rho}(Z_\theta, \gamma)}{\partial \gamma} \bigg|_{\theta = \theta^0, \gamma = 0} \right] = \Gamma_V
\]

\[
E \left[ \frac{\partial g_{\theta, \rho}(Z_\theta, \gamma)}{\partial \gamma'} \bigg|_{\theta = \theta^0, \gamma = 0} \right] = \Sigma_E
\]

Define the following notations:

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{\theta, \rho}(Z_t, \theta, \gamma)}{\partial \gamma'} \bigg|_{\theta = \hat{\theta}, \gamma = \hat{\gamma}} = \hat{\Gamma}'_{V,T}
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{\theta, \rho}(Z_t, \theta, \gamma)}{\partial \gamma'} \bigg|_{\theta = \hat{\theta}, \gamma = \hat{\gamma}} = \hat{\Sigma}'_{E,T}
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{\theta, \rho}(Z_t, \theta, \gamma)}{\partial \gamma} = \hat{\Gamma}'_{V,T}(\theta, \gamma)
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{\theta, \rho}(Z_t, \theta, \gamma)}{\partial \gamma'} = \hat{\Sigma}'_{E,T}(\theta, \gamma)
\]

Because \( \theta^0 \in \text{interior}(\Theta) \) and \( \tilde{g}(Z, \theta, \gamma) \) is continuously differentiable in a neighborhood \( N \) of \( (\theta^0, G^0) \), with probability approaching one the first-order conditions

\[
\hat{\Sigma}'_{E,T} \hat{W} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{\theta, \rho}(z_t, \hat{\theta}, \hat{\gamma}) = 0
\]

\[
\hat{\Gamma}'_{V,T} \hat{W} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{\theta, \rho}(z_t, \hat{\theta}, \hat{\gamma}) = 0
\]

are satisfied. Expanding \( \sum_{t=1}^{T} g_{\theta, \rho}(z_t, \hat{\theta}, \hat{\gamma}) \) around \( (\theta^0, 0) \) gives

\[
\begin{pmatrix}
\hat{\Sigma}'_{E,T} \hat{W} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0) \\
\hat{\Gamma}'_{V,T} \hat{W} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0)
\end{pmatrix}
\begin{pmatrix}
\sqrt{T} \hat{\gamma} \\
\sqrt{T}(\hat{\theta} - \theta^0)
\end{pmatrix}
\]

Where a bar denotes a mean value. Since the expansion is around a fixed but essentially arbitrary point in \( G^0 \) and I cannot expect to have \( \hat{\gamma}_T \rightarrow 0 \), the appearance of the mean values \( \bar{\gamma} \) seems to
require some additional care. I deduce now that the distance between these mean values and the set $G^0$ converges in probability to zero. Recall that by Lemma 2.2.2 that $G^0$ is convex. For any $\epsilon > 0$ let $G_\epsilon = \{ \gamma + c \in \mathbb{R}^H | \gamma \in G^0, \|c\| < \epsilon \}$ denote the epsilon fattening of $G^0$. Convexity of $G_\epsilon$ is implied by the convexity of $G^0$. Since the mean value belongs to the line segment that connects $\hat{\gamma}$ and the origin the convexity of $G_\epsilon$ means that if $\hat{\gamma} \in G_\epsilon$ then $\bar{\gamma} \in G_\epsilon$. By Theorem 2.2.1, $\hat{\gamma} \in G_\epsilon$ with probability approaching one for any $\epsilon > 0$, so that $\bar{\gamma} \in G_\epsilon$ with probability approaching one for any $\epsilon > 0$ as well. Therefore, $d(\Gamma, \bar{\gamma}) \rightarrow p 0$ and

$$
\frac{d\rho(u|z; \hat{\gamma}, \hat{\theta})}{d\rho(\cdot|z; \theta)} \rightarrow_p \frac{d\rho(u|z; 0, \theta^0)}{d\rho(\cdot|z; \theta^0)}.
$$

This result combined with the integrability conditions imply that the sample derivatives have the expected limits and

$$
\begin{bmatrix}
\Sigma_E(\Sigma_V^{L,R}) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0) \\
\Gamma_V(\Sigma_V^{L,R}) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0)
\end{bmatrix} + o_p(1) = (H + o_p(1)) \begin{bmatrix}
\sqrt{T} \hat{\gamma} \\
\sqrt{T}(\hat{\theta} - \theta^0)
\end{bmatrix}
$$

(2.21)

where:

$$
H = \begin{bmatrix}
\Sigma_E(\Sigma_V^{L,R}) + \Sigma_E & 0 \\
0 & \Gamma_V(\Sigma_V^{L,R}) + \Gamma_V
\end{bmatrix}
$$

Thanks to the assumption that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0)$ satisfies a central limit theorem the left side of (2.21) has an asymptotic distribution given by:

$$
\begin{bmatrix}
\Sigma_E(\Sigma_V^{L,R}) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0) \\
\Gamma_V(\Sigma_V^{L,R}) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta^0}(Z_t, \theta^0)
\end{bmatrix} \rightarrow^d N(0, H)
$$

The asymptotic distribution of the right side of (2.21) is also:

$$
\begin{bmatrix}
\Sigma_E(\Sigma_V^{L,R}) + \Sigma_E \sqrt{T} \hat{\gamma} \\
\Gamma_V(\Sigma_V^{L,R}) + \Gamma_V \sqrt{T}(\hat{\theta} - \theta^0)
\end{bmatrix} \rightarrow^d N(0, H)
$$
Therefore:

$$\Gamma'_{V}(\Sigma^{LR}_{V})+\Gamma_{V}\sqrt{T}(\hat{\theta} - \theta^{0}) \rightarrow^{d} N(0, \Gamma'_{V}(\Sigma^{LR}_{V})+\Gamma_{V})$$

Finally noting that $\Gamma'_{V}(\Sigma^{LR}_{V})+\Gamma_{V}$ is invertible by Assumption 2.2.5:

$$\sqrt{T}(\hat{\theta} - \theta^{0}) \rightarrow^{d} N(0, (\Gamma'_{V}(\Sigma^{LR}_{V})+\Gamma_{V})^{-1})$$

\[\square\]

**Proof of Theorem 2.2.3:**

*Proof.* Let

$$S = \begin{pmatrix} Id & -W_{12}W_{22}^{-1} \\ 0 & Id \end{pmatrix} B$$

and without loss of generality assume that the reference measure is given by the measure $\mu^{S}$ from lemma 2.8.2 by Assumption 2.2.2. Following the same reasoning as in the proof of Theorem 2.2.2:

$$\begin{pmatrix} \Sigma'_{E}W\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{g}_{\theta^{0}}(Z_{t},\theta^{0}) \\ \Gamma'_{V}W\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{g}_{\theta^{0}}(Z_{t},\theta^{0}) \end{pmatrix} + o_{p}(1) = (J + o_{p}(1)) \begin{pmatrix} \sqrt{T}\tilde{\gamma} \\ \sqrt{T}(\hat{\theta} - \theta^{0}) \end{pmatrix}$$

with:

$$J = \begin{pmatrix} \Sigma_{E}W\Sigma_{E} & 0 \\ 0 & \Gamma'_{V}WT_{V} \end{pmatrix}$$

So by Assumption 2.2.6:

$$\sqrt{T}(\hat{\theta} - \theta^{0}) \rightarrow^{d} N(0, (\Gamma'_{V}(\Sigma^{LR}_{V})+\Gamma_{V})^{-1})$$
Thanks to the assumption that \( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta_0}(Z_t, \theta_0) \) satisfies a central limit theorem:

\[
\Gamma'_V W \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\theta_0}(Z_t, \theta_0) \rightarrow^d N(0, \Gamma'_V W \Sigma^L R W \Gamma_V)
\]

So the asymptotic distribution of the right side of (2.22) is also:

\[
(\Gamma'_V W T \sqrt{T}(\hat{\theta} - \theta^0) \rightarrow^d N(0, \Gamma'_V W \Sigma^L R W \Gamma_V)
\]

Noting that \( \Gamma'_V W T \) is invertible by Assumption 2.2.6:

\[
\sqrt{T}(\hat{\theta} - \theta^0) \rightarrow^d N \left( 0, (\Gamma'_V (\Sigma^L R)^+ \Gamma_V)^{-1} \Gamma'_V W \Sigma^L R W \Gamma_V (\Gamma'_V (\Sigma^L R)^+ \Gamma_V)^{-1} \right)
\]

Finally:

\[
\Gamma'_V (\Sigma^L R)^+ \Gamma_V = (ST_V)'(S^{-1} W S^{-1}) (ST_V)
\]

\[
\Rightarrow \Gamma'_V (\Sigma^L R)^+ \Gamma_V = E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]' W^{-1} \Sigma_22 W^{-1} E \left[ \frac{\partial}{\partial \theta} h_2(U_t, Z_t, \theta)|_{\theta = \theta^0} \right]
\]

**Proof of Theorem 2.5.1:**

*Proof.* The first order conditions of the Lagrangian (2.8) are given by

\[
\varphi_{\pi}(\hat{\pi}_t(\hat{\theta}, \hat{\gamma})) = \hat{\beta}'_T \Psi_{t, T, \rho}(\hat{\theta}, \hat{\gamma}) + \mu \tag{2.23}
\]

\[
\hat{\beta}'_T \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \Psi_{t, T, \rho}}{\partial \theta}(\hat{\theta}, \hat{\gamma}) = 0 \tag{2.24}
\]

\[
\hat{\beta}'_T \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \Psi_{t, T, \rho}}{\partial \gamma}(\hat{\theta}, \hat{\gamma}) = 0 \tag{2.25}
\]
Right multiply (2.23) by \( \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})' \) and sum over \( t \) to obtain

\[
\sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\varphi_\pi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma}))\Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})' = \hat{\beta}'_T \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})' = \hat{\beta}'_T \text{Var}_\pi [E_{\rho}(g(U, Z, \theta)|Z)]
\]

Right multiply by \( \hat{N}_T(\hat{\theta}, \hat{\gamma})' \) to obtain

\[
\sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\varphi_\pi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma}))(\hat{N}_T(\hat{\theta}, \hat{\gamma})\Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})')' = (\hat{N}_T(\hat{\theta}, \hat{\gamma})^{-1}'\hat{\beta}_T)' \text{Var}_\pi [E_{\rho}(\hat{N}_T(\hat{\theta}, \hat{\gamma})g(U, Z, \theta)|Z)]
\]

(2.26)

and rewrite (2.26) as

\[
\sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\varphi_\pi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma}))\Psi_{t,T,\rho,1}(\hat{\theta}, \hat{\gamma})', \Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma})' = (\hat{\beta}'_{T,1}, \hat{\beta}'_{T,2}) \begin{pmatrix} \hat{\Sigma}_{11} & 0 \\ 0 & \hat{\Sigma}_{22} \end{pmatrix}
\]

from which I deduce:

\[
\hat{\beta}'_{T,2} = \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma})\varphi_\pi(\hat{\pi}_t(\hat{\theta}, \hat{\gamma}))\Psi_{t,T,\rho,2}(\hat{\theta}, \hat{\gamma})' \hat{\Sigma}_{22}^{-1}
\]

(2.27)

Right multiply (2.25) by \( \hat{N}_T(\hat{\theta}, \hat{\gamma})' \) to obtain

\[
(\hat{N}_T(\hat{\theta}, \hat{\gamma})^{-1}'\hat{\beta}_T)'\hat{N}_T(\hat{\theta}, \hat{\gamma}) \sum_{t=1}^{T} \hat{\pi}_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \Psi_{t,T,\rho}(\hat{\theta}, \hat{\gamma})}{\partial \gamma'} (\hat{\theta}, \hat{\gamma}) \hat{N}_T(\hat{\theta}, \hat{\gamma})' = 0
\]

which can be rewritten as

\[
(\hat{\beta}'_{T,1}, \hat{\beta}'_{T,2}) E_\pi [\text{Var}_\rho(\hat{N}_T(\theta, \gamma)g(U, Z, \theta)|Z)] = 0
\]

and using (2.17)

\[
(\hat{\beta}'_{T,1}, \hat{\beta}'_{T,2}) \begin{pmatrix} \hat{\Lambda} & 0 \\ 0 & 0 \end{pmatrix} = 0
\]
from which I deduce:

$$\hat{\beta}_{T,1}^t = 0$$  \hspace{1cm} (2.28)

Finally, substitute (2.27) and (2.28) into the left side of (2.24) as

$$\hat{\beta}_{T}^t \sum_{t=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \Psi_{T, k}(\hat{\theta}, \hat{\gamma})}{\partial \theta} = (\hat{\beta}_{T,1}^t)^{-1} \hat{\beta}_{T}^t \sum_{t=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \hat{\beta}_{T}^t(\hat{\theta}, \hat{\gamma})}{\partial \theta}$$

Taking the transpose of the previous display yields the result:

$$\left( \sum_{t=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \psi_{T, k}(\hat{\theta}, \hat{\gamma})}{\partial \theta} \right) \left( \sum_{t=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \psi_{T, k, 2}(\hat{\theta}, \hat{\gamma})}{\partial \theta} \right)^{-1} \left( \sum_{t=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \psi_{T, k, 2}(\hat{\theta}, \hat{\gamma})}{\partial \theta} \right) = 0$$

Proof of Theorem 2.6.1

Proof. The first order conditions of the Lagrangian are given by

$$w(t, j) \frac{\partial \psi_{T, k}(\hat{\theta}, \hat{\gamma})}{\partial \theta} = (\hat{\beta}_{T,1}^t)^{-1} (\hat{\beta}_{T,2}^t) \frac{\partial \psi_{T, k, 2}(\hat{\theta}, \hat{\gamma})}{\partial \theta}$$

$$\sum_{t=1}^{T} \hat{\beta}_{T,1}^t \sum_{j=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \psi_{T, k}(\hat{\theta}, \hat{\gamma})}{\partial \theta} = 0$$

$$\hat{\beta}_{T,1}^t \sum_{j=1}^{T} \pi_t(\hat{\theta}, \hat{\gamma}) \frac{\partial \psi_{T, k, 2}(\hat{\theta}, \hat{\gamma})}{\partial \theta} = 0 \hspace{1cm} \forall t \in \{1, \ldots, n\}$$
Right multiply (2.29) by \( \hat{\pi}_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)' \) and sum over \( j \) to obtain

\[
\sum_{j=1}^{T} w_{tj}^\lambda \pi_{tj}(\hat{\theta}, \hat{\gamma}_t)^{1-\lambda} \Psi_j(\hat{\theta}, \hat{\gamma}_t)' = \beta_t' \sum_{j=1}^{T} \hat{\pi}_{tj}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)' = \beta_t'[\text{Var}_\pi[E_{n_i}(g(U_t, Z_t, \theta)|Z_t, Z_{t-1})|Z_{t-1} = z_{t-1}]]
\]

Right multiply by \( \hat{N}_T(\hat{\theta}, \hat{\gamma}_t)^{t-1}' \) to obtain

\[
\sum_{j=1}^{T} w_{tj}^\lambda \pi_{tj}(\hat{\theta}, \hat{\gamma}_t)^{1-\lambda}(\hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t) \Psi_j(\hat{\theta}, \hat{\gamma}_t)') = (\hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t)^{-1}' \beta_t)' \text{Var}_\pi[E_{n_i}(\hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t)g(U_t, Z_t, \theta)|Z_t, Z_{t-1})|Z_{t-1} = z_{t-1}] \tag{2.32}
\]

and rewrite (2.32) as

\[
\sum_{j=1}^{T} w_{tj}^\lambda \pi_{tj}(\hat{\theta}, \hat{\gamma}_t)^{1-\lambda}(\Psi_{j,1}(\hat{\theta}, \hat{\gamma}_t)', \Psi_{j,2}(\hat{\theta}, \hat{\gamma}_t)') = (\beta_{t,1}', \beta_{t,2}') \left( \begin{array}{c|c}
\hat{\Sigma}^{t}_{11} & 0 \\
0 & \hat{\Sigma}^{t}_{22} 
\end{array} \right) \tag{2.33}
\]

from which I deduce:

\[
\beta_{t,2}' = \sum_{j=1}^{T} w_{tj}^\lambda \pi_{tj}(\hat{\theta}, \hat{\gamma}_t)^{1-\lambda}(\Psi_{j,2}(\hat{\theta}, \hat{\gamma}_t)') (\hat{\Sigma}^{t-1}_{22})^{-1} \tag{2.34}
\]

Right multiply (2.31) by \( \hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t)' \) to obtain:

\[
\sum_{t=1}^{T} \beta_t' \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial \Psi_j}{\partial \gamma'_t}(\hat{\theta}, \hat{\gamma}_t) \hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t)' = 0
\]

Now, rewriting:

\[
\sum_{t=1}^{T} (\hat{N}_T^{t}(\hat{\theta}, \hat{\gamma}_t)^{-1}' \beta_t') \hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t) \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial \Psi_j}{\partial \gamma'_t}(\hat{\theta}, \hat{\gamma}_t) \hat{N}_T^{t-1}(\hat{\theta}, \hat{\gamma}_t)' = 0
\]

Therefore:

\[
\sum_{t=1}^{T} (\beta_{t,1}', \beta_{t,2}') E_{\pi}[\text{Var}_{n_i}(\hat{N}_T^{t-1}(\theta, \gamma_t)g(U_t, Z_t, \theta)|Z_t, Z_{t-1})|Z_{t-1} = z_{t-1}] = 0
\]
And using (2.17):

\[
\sum_{t=1}^{T} (\hat{\beta}'_{t,1}, \hat{\beta}'_{t,2}) \left( \frac{\hat{\Lambda}^{-1} - 0}{0} \right) = 0
\]

Because \( \hat{\Lambda}^{-1} \) is nonsingular and the Lagrange multipliers are nonnegative, I deduce:

\[
\hat{\beta}'_{t,1} = 0 \quad \forall t \in \{1, \ldots, T\}
\]  

(2.35)

Finally, substitute (2.34) and (2.35) into the left side of (2.30) as

\[
\sum_{t=1}^{T} \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial \psi_j(\hat{\theta}, \hat{\gamma}_t)}{\partial \theta'}
\]

\[
= \sum_{t=1}^{T} \left( \sum_{j=1}^{T} w_{\gamma}^{\lambda} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) 1 - \lambda \psi_j,2(\hat{\theta}, \hat{\gamma}_t) (\hat{\Sigma}_{22}^{-1})^{-1} \right) \left( \sum_{j=1}^{T} \frac{\partial \psi_j,1(\hat{\theta}, \hat{\gamma}_t)}{\partial \theta'} \psi_{j,2}(\hat{\theta}, \hat{\gamma}_t) \right)
\]

Taking the transpose of the previous display yields the result:

\[
\sum_{t=1}^{T} \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \frac{\partial \psi_j,1(\hat{\theta}, \hat{\gamma}_t)}{\partial \theta'} \right) \left( \sum_{j=1}^{T} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) \psi_{j,2}(\hat{\theta}, \hat{\gamma}_t) \right)^{-1} \left( \sum_{j=1}^{T} w_{\gamma}^{\lambda} \pi_{tj}(\hat{\theta}, \hat{\gamma}_t) 1 - \lambda \psi_j,2(\hat{\theta}, \hat{\gamma}_t) \right) = 0
\]

2.9 Appendix: Figures and Tables
Figure 2.1: Monte Carlo sampling distributions for the two-step efficient ELVIS estimator and the minimum distance estimator of Gospodinov et al. (2017) with $\beta = 1, 1/4$. 
Figure 2.2: Monte Carlo sampling distributions for the two-step efficient ELVIS estimator and the minimum distance estimator of Gospodinov et al. (2017) with $\beta = 1/8, 0$. 
Table 2.1: Monte Carlo results for the two-step efficient ELVIS estimator and the minimum distance estimator of Gospodinov et al. (2017).

<table>
<thead>
<tr>
<th>β = 1</th>
<th>ELVIS</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9880</td>
<td>1.0294</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.2600</td>
<td>0.2849</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0678</td>
<td>0.0820</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β = 1/4</th>
<th>ELVIS</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.2806</td>
<td>-0.0100</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.3119</td>
<td>2.8712</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0982</td>
<td>8.3112</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β = 1/8</th>
<th>ELVIS</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.1486</td>
<td>0.3999</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.1769</td>
<td>2.0554</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0319</td>
<td>4.3004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β = 0</th>
<th>ELVIS</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0014</td>
<td>0.0514</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.1155</td>
<td>2.4655</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0133</td>
<td>6.0812</td>
</tr>
</tbody>
</table>

Table 2.2: Estimation Results for the two-step efficient ELVIS estimator, the minimum distance estimator of Gospodinov et al. (2017) and the indirect inference estimator of Grammig and Küchlin (2018).

<table>
<thead>
<tr>
<th>ELVIS</th>
<th>MD</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point Estimate</td>
<td>4.11</td>
<td>2.57</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.60</td>
<td>1.18</td>
</tr>
<tr>
<td>Lower</td>
<td>2.93</td>
<td>0.26</td>
</tr>
<tr>
<td>Upper</td>
<td>5.29</td>
<td>4.88</td>
</tr>
</tbody>
</table>
Chapter 3

Efficient Errors-in-Variables with Multiple Proxies

3.1 Introduction

Measurement error in economic variables and the resulting bias in structural parameters estimates is a pervasive issue in applied econometrics. A good deal of attention has been paid to the case of regressions with multiple proxies (see Lubotsky and Wittenberg 2006 and references therein) where the attenuation bias of OLS can be reduced if more information is available through the observation of additional proxies of the variables measured with errors. As soon as two measures of an error-ridden variable are available, one can solve the errors-in-variable problem by using one measure as an instrument for the other. However, as stressed by Andersson and Møen (2016), a second estimate can be produced by reversing the role of the variable and the instrument and while “several studies present both estimates, (…) no discussion of the choice between them appears to be available in the econometrics literature”. The contribution of Andersson and Møen (2016) was to characterize the optimal linear combination of these two estimates.

However, since IV estimation is a particular case of the Generalized Method of Moments (GMM), it is natural to think about characterization of an optimal use of the available information by applying the theory of efficient GMM of Hansen (1982). This is the attempt made by Chalfin and McCrary (2017) in the context of the afore-described classical errors-in-variable model. Even though they
do not refer to Andersson and Møen (2016), they claim that "practically, to pool the forward and reflected IV estimates, we stack the orthogonality conditions for the forward and the reflected IV programs into the broader set of moments (...)". Note that this claim does not give an explicit answer to two important questions:

First, is it equivalent to the optimal way "to pool the forward and reflected IV estimates" as characterized by Andersson and Møen (2016)? We actually know from a particular case of Chen et al. (2016) that the two approaches should deliver asymptotically equivalent estimators.

Second, and even more importantly, is it the most efficient GMM estimator that can be deduced from the maintained assumptions of this regression model with multiple proxies? This chapter can be seen as an answer to this efficiency issue put forward by Chalfin and McCrary (2017).

This chapter is organized as follows. The errors-in-variables model with multiple proxies and various estimation strategies are introduced in section 3.2. Section 3.3 presents analytical and simulation based results for a simple bivariate version of errors-in-variables model. Section 3.4 gives the main result of this chapter which characterizes when the estimators of Andersson and Møen (2016) and Chalfin and McCrary (2017) attain the semiparametric efficiency bound. Section 3.5 concludes. Proofs appear in the appendix.

### 3.2 Model and Estimation

In this chapter I consider an affine classical measurement error model. The regressor $X^*$ is latent, but two proxies $X_1$ and $X_2$ are observed in addition to $k$ covariates $\xi$:

\[
Y = \xi'\beta + X^*\gamma + \Delta Y \\
X_1 = X^* + \Delta X_1 \\
X_2 = X^* + \Delta X_2.
\]

The observables in this model are $Z = (Y, X_1, X_2, \xi')'$, the unobservables are $U = X^*$, and the parameter vector is $\theta = (\gamma, \beta')'$. Following Chalfin and McCrary (2017) I assume that the sample
data is independent and identically distributed and maintain the following assumptions:

\[ E[\Delta X_1 \Delta Y] = E[\Delta X_2 \Delta Y] = 0 \]
\[ E[\Delta X_1 (X^*, \xi')] = E[\Delta X_2 (X^*, \xi')] = 0 \]
\[ E[\Delta X_1 \Delta X_2] = 0 \]
\[ E[\Delta Y (X^*, \xi')] = 0. \]

I will consider four ways to estimate this model.

### 3.2.1 Instrument Variables Estimation

With two proxies available the researcher, IV approaches to parameter estimation are available. Letting \( Z'_1 = (\xi', X_1) \), and \( Z'_2 = (\xi', X_2) \) there are two possible IV estimators obtained from the sample analog of the moment restrictions:

\[ E[(Y - Z'_1 \theta) Z_2] = 0 \]  
(3.1)
\[ E[(Y - Z'_2 \theta) Z_1] = 0 \]  
(3.2)

I denote the IV estimate obtained from (3.1) as \( \hat{\theta}^{IV}_1 \) and (3.2) as \( \hat{\theta}^{IV}_2 \). These IV estimates are, in general, not efficient. Intuitively, these estimators each leave some information about the parameter unexploited and combining the information in (3.1) and (3.2) will result a reduced asymptotic variance.

### 3.2.2 Andersson and Møen (2016) Estimation

The approach of Andersson and Møen (2016) combines the two IV estimators of the previous section as:

\[ \hat{\theta}^{AM} = w_1 \hat{\theta}^{IV}_1 + w_2 \hat{\theta}^{IV}_2 \]
where the weights are chosen to minimize asymptotic variance. The optimal weights are given by

\[
w_1 = \frac{\text{Var}(\hat{\theta}^{IV}_2) - \text{Cov}(\hat{\theta}^{IV}_1, \hat{\theta}^{IV}_2)}{\text{Var}(\hat{\theta}^{IV}_1) + \text{Var}(\hat{\theta}^{IV}_2) - 2\text{Cov}(\hat{\theta}^{IV}_1, \hat{\theta}^{IV}_2)}
\]

\[
w_2 = 1 - w_1
\]

and are estimated using sample analogs.

### 3.2.3 Chalfin and McCrary (2017) Estimation

The approach of Chalfin and McCrary (2017) is meant to combine the two IV estimators implicitly by using the moment restrictions corresponding to each IV estimator in a GMM estimator. The moment restrictions for the Chalfin McCrary estimator are the same as (3.1) and (3.2) above:

\[
g_{CM}(Z, \theta) = \begin{pmatrix}
(Y - \xi'\beta - X_2\gamma)X_1 \\
(Y - \xi'\beta - X_1\gamma)X_2 \\
(Y - \xi'\beta - X_1\gamma)\xi \\
(Y - \xi'\beta - X_2\gamma)\xi
\end{pmatrix}
\]

(3.3)

I will denote by \(\hat{\theta}_{CM}\) the estimator obtained from GMM estimation using these moment restrictions and an optimal weighting matrix. As noted in above, from the results of Chen et al. (2016) \(\hat{\theta}_{CM}\) and \(\hat{\theta}_{AM}\) are asymptotically equivalent. A drawback of the Chalfin McCrary estimator is that if \(\gamma = 0\), the set of moment restrictions become linearly dependent so that the variance covariance matrix of the moment is singular.
3.2.4 ELVIS Estimation

The framework of Chapters 1 and 2 may be applied to this model. The assumptions imply the following moment restrictions:

\[
g(U, Z, \theta) = \begin{pmatrix}
(Y - \xi' \beta - X^* \gamma) \Delta X_1 \\
(Y - \xi' \beta - X^* \gamma) \Delta X_2 \\
\xi \Delta X_1 \\
\xi \Delta X_2 \\
\Delta X_1 \\
\Delta X_2 \\
\Delta X_1 \Delta X_2 \\
(Y - \xi' \beta - X^* \gamma) X^* \\
(Y - \xi' \beta - X^* \gamma) \xi \\
\Delta X_1 X^* \\
\Delta X_2 X^*
\end{pmatrix}
\]

The model falls within the framework of Chapters 1 and 2 with \( Z = (Y, \xi', X_1, X_2) \) observable and \( U = X^* \) latent.

3.3 Example: bivariate model

To fix intuition I consider a bivariate classical measurement error model where two mismeasurements of the regressor are observed and explicitly compute the asymptotic variance for each estimator. In addition, the data generating process is fully specified to allow for explicit calculation of efficiency bound. The model is given by

\[Y = X^* \theta + \Delta Y\]

Where the observables are:

\[Y, \quad X_1 = X^* + \Delta X_1, \quad X_2 = X^* + \Delta X_2\]
and $\Delta X_1, \Delta X_2$, and $\Delta Y$ are mean zero and uncorrelated from one another and $X^*$. This model is equivalent to the set of 9 moment restrictions:

$$g(U, Z, \theta) = \begin{pmatrix} \Delta X_1 \\ \Delta X_2 \\ \Delta Y \\ X^*\Delta X_1 \\ X^*\Delta X_2 \\ X^*\Delta Y \\ \Delta Y\Delta X_1 \\ \Delta Y\Delta X_2 \\ \Delta X_1\Delta X_2 \end{pmatrix} = \begin{pmatrix} X_1 - X^* \\ X_2 - X^* \\ Y - X^*\theta \\ X^*(X_1 - X^*) \\ X^*(X_2 - X^*) \\ X^*(Y - X^*\theta) \\ (Y - X^*\theta)(X_1 - X^*) \\ (Y - X^*\theta)(X_2 - X^*) \\ (X_1 - X^*)(X_2 - X^*) \end{pmatrix}.$$ 

The model falls within the framework of Chapters 1 and 2 with $Z = (Y, X_1, X_2)$ observable and $U = X^*$ latent. From the discussion in Chapter 1, the efficiency bound is given by the “important directions” of the moment function which are characterized by the spectral decomposition of $\Sigma_E = E[Var(g(U, Z, \theta^0)|X_1, X_2, Y)]$. The following proposition uses the results of Chapter 1 to explicitly compute the efficiency bound for this model.

**Proposition 3.3.1.** Suppose that the true DGP is given by:

$$(\Delta X, \Delta Y, \Delta Z, X^*) \sim N(0, I_4) \quad \theta^0 = 0.$$ 

The efficiency bound for the estimation of $\theta^0$ in the bivariate linear model is $\frac{3}{2}$. 

From the results of Chapter 2, an ELVIS estimator based on the moment restrictions above will have asymptotic variance equal to $\frac{3}{2}$. As expected, the IV estimates are not efficient and standard calculations show each have asymptotic variance equal to 2. The covariance between the two IV estimators is equal to 1 so that the Andersson Møen weights are equal to $\frac{1}{2}$ and the asymptotic
variance of $\hat{\theta}^{AM}$ is $\frac{3}{2}$. The Chalfin McCrary estimator is formed using the moment restrictions:

$$g^{CM}(Z, \theta) = \begin{pmatrix} (Y - X_2 \theta)X_1 \\ (Y - X_1 \theta)X_2 \end{pmatrix}.$$

So that the optimal weighting matrix and Jacobian are given by:

$$Var[g^{CM}(Z, \theta^0)]^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix}$$

$$E \left[ \frac{g^{CM}(Z, \theta^0)}{\partial \theta} \right] = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So the asymptotic variance is given by:

$$\left( E \left[ \frac{g^{CM}(Z, \theta^0)}{\partial \theta} \right] \right)' Var[g^{CM}(Z, \theta^0)]^{-1} E \left[ \frac{g^{CM}(Z, \theta^0)}{\partial \theta} \right]^{-1} = \frac{3}{2}.$$

Therefore, in this simple case, with no covariates and Gaussian data generating process, the Andersson Møen and Chalfin McCrary estimators are asymptotically efficient. The next section considers general conditions under which these estimators attain an asymptotic efficiency bound.

### 3.3.1 Monte Carlos Results

In this section I present some Monte Carlo results from the simple bivariate model discussed above. Replications are simulated from the Gaussian data generating process:

$$(\Delta X, \Delta Y, \Delta Z, X^*) \sim N(0, I_4)$$

with $\theta^0 = 0$ for sample sizes $n = 200$ and $n = 500$. For each replication I compute the two-step efficient ELVIS, Andersson Møen and Chalfin McCrary estimators.

I simulate 5000 replications for each sample size. The mean, standard deviation, and mean square error for each estimator are reported in table 3.1. Figure 3.1 contains a density plot of the results for each estimator when $n = 200$ and figure 3.2 contains a density plot of the results for each estimator when $n = 500$. 

As anticipated by the analytic results presented above, each estimator performs similarly. For both \( n = 200 \) and \( n = 500 \), the sampling distributions appear to approximately Gaussian with MSE close to the theoretical predictions of 0.0075 and 0.003 respectively.

### 3.4 General Case

In this section I consider the general case of a regression model with a vector of covariates, \( \xi \), and two proxies. Recall from section 3.2, the model is given by

\[
Y = \xi' \beta + \gamma^* + \Delta Y
\]

Where the observables are:

\[
Y, \quad \xi, \quad X_1 = X^* + \Delta X_1, \quad X_2 = X^* + \Delta X_2
\]

where \( \Delta Y, \Delta X_1, \Delta X_2 \) are mean zero and mutually uncorrelated.

The following theorems give conditions under which the Chalfin McCrary estimator will attain the semiparametric efficiency bound of chapter 1. The required conditions vary slightly whether the coefficient on the mismeasured regressor \( \gamma^0 \) is zero or nonzero. Theorem 3.4.1 addresses the case where \( \gamma^0 \neq 0 \) and Theorem 3.4.2 the case where \( \gamma^0 = 0 \).

In either case, a required condition is that the difference of the proxies \( X_1, X_2 \) is uncorrelated from the from the Chalfin McCrary moment restrictions given in equation (3.3). The equivalence of the three estimators in section 3.3 is not surprising as the Gaussian data generating process implies that the third order moments arising in these correlations are equal to zero.
Theorem 3.4.1. Suppose $\gamma^0 \neq 0$. The Chalfin McCrary estimator is efficient if

$$\det[E[Var(\begin{pmatrix} \Delta X_2 \\ \Delta X_1 \Delta X_2 \\ (Y - \xi'\beta - X^*\gamma)X^* \\ (Y - \xi'\beta - X^*\gamma)\xi \\ \Delta X_1 X^* \\ \Delta X_2 X^* \end{pmatrix}|Z])] \neq 0$$

and $\text{Cov}[g^{CM}(Z, \theta^0), X_1 - X_2] = 0$.

Theorem 3.4.2. Suppose $\gamma^0 = 0$. The Chalfin McCrary estimator is efficient if

$$\det[E[Var(\begin{pmatrix} \xi \Delta X_1 \\ \xi \Delta X_2 \\ \Delta X_2 \\ \Delta X_1 \Delta X_2 \\ (Y - \xi'\beta^0)X^* \\ \Delta X_1 X^* \\ \Delta X_2 X^* \end{pmatrix}|Z])] \neq 0$$

and $\text{Cov}[g^{CM}(Z, \theta^0), X_1 - X_2] = 0$.

3.5 Conclusion

In this chapter I have examined the issue of semiparametric efficiency in classical errors-in-variables models when two proxies are available. I have shown, analytically and through Monte Carlo experiments the efficiency properties of the estimation strategies put forward in Andersson and Møen (2016) and Chalfin and McCrary (2017). Further theoretical research would examine the case of three or more proxies.
3.6 Appendix: Proofs

Proof of Proposition 3.3.1

Proof. Under the assumption of normality it is possible to explicitly compute the function \( h_2(U, Z, \theta^0) \) to find the efficiency bound. Recall that the function \( h_2(U, Z, \theta^0) \) is constructed from the spectral decomposition of \( \Sigma_E = E[Var(g(U, Z, \theta^0)|X_1, X_2, Y)] \). To compute \( \Sigma_E \) note that:

\[
X^*|X_1, X_2, Y \sim N \left( \frac{X_1 + X_2}{3}, \frac{1}{3} \right).
\]

Therefore, the following identities arise from standard calculations:

\[
E[X^*|X_1, X_2, Y] = \frac{X_1 + X_2}{3},
\]

\[
E[X^*^2|X_1, X_2, Y] = \frac{1}{3} + \left( \frac{X_1 + X_2}{3} \right)^2,
\]

\[
E[X^*^3|X_1, X_2, Y] = \left( \frac{X_1 + X_2}{3} \right)^3 + \frac{X_1 + X_2}{3},
\]

\[
E[X^*^4|X_1, X_2, Y] = \left( \frac{X_1 + X_2}{3} \right)^4 + 2\left( \frac{X_1 + X_2}{3} \right)^2 + \frac{1}{3},
\]

\[
Cov[X^*^2, X^*|X_1, X_2, Y] = \frac{2}{9}(X_1 + X_2).
\]

With these results in hand the conditional variance matrix is given by:

\[
\Sigma_E = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/9 & 1/9 & 0 & 0 & 0 & -1/9 & \\
0 & 0 & 0 & 1/9 & 1/9 & 0 & 0 & 0 & -1/9 & \\
0 & 0 & 0 & 0 & 1/3 & -1/3 & -1/3 & 0 & \\
0 & 0 & 0 & 0 & -1/3 & 1/3 & 1/3 & 0 & \\
0 & 0 & 0 & 0 & -1/3 & 1/3 & 1/3 & 0 & \\
0 & 0 & 0 & -1/9 & -1/9 & 0 & 0 & 0 & 4/9
\end{pmatrix}.
\]
The Eigenvalues of $\Sigma_E$ are

$$
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
1/3 + 1/9 \sqrt{3} \\
1/3 - 1/9 \sqrt{3} \\
2/3 \\
1
\end{bmatrix}
$$

So the function $g_2(U, Z, \theta)$ will have 5 components. Five orthonormal eigenvectors for the characteristic value zero are given by

$$
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
-\sqrt{2}/\sqrt{6} \\
\sqrt{2}/\sqrt{6} \\
0 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
\sqrt{2}/\sqrt{2} \\
-\sqrt{2}/\sqrt{2} \\
0 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
0 \\
1 \\
0 \\
0 \\
-\sqrt{2}/\sqrt{6} \\
\sqrt{2}/\sqrt{6} \\
0 \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\sqrt{2}/\sqrt{2} \\
-\sqrt{2}/\sqrt{2} \\
0 \\
0
\end{bmatrix},
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\sqrt{2}/\sqrt{6} \\
\sqrt{2}/\sqrt{6} \\
0
\end{bmatrix}
$$
Thus,

\[
h_2(U, Z, \theta) = \begin{pmatrix}
0 & 0 & 0 & 0 & -\sqrt{6} & \sqrt{6} & -\sqrt{6} & 0 \\
-\sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
X_1 - X^* \\
X_2 - X^* \\
Y - X^* \theta \\
X^*(X_1 - X^*) \\
X^*(X_2 - X^*) \\
X^*(Y - X^* \theta) \\
(Y - X^* \theta)(X_1 - X^*) \\
(Y - X^* \theta)(X_2 - X^*) \\
(X_1 - X^*)(X_2 - X^*)
\end{pmatrix}.
\]

Therefore,

\[
E \left[ \frac{\partial h_2(U, Z, \theta)}{\partial \theta} \Bigg|_{\theta = \theta_0} \right] = \begin{pmatrix}
-\sqrt{6}/6 \\
0 \\
0 \\
0 \\
\sqrt{2}/2
\end{pmatrix}
\]

Finally,

\[
\left( E \left[ \frac{\partial h_2(U_t, Z_t, \theta)}{\partial \theta} \Bigg|_{\theta = \theta_0} \right] \right) \Sigma_{22}^{-1} E \left[ \frac{\partial h_2(U_t, Z_t, \theta)}{\partial \theta} \Bigg|_{\theta = \theta_0} \right]^{-1} = \begin{pmatrix}
-\sqrt{6}/6, 0, 0, 0, \sqrt{2}/2
\end{pmatrix}^{-1}
\]

\[
= \begin{pmatrix}
3/2
\end{pmatrix}
\]

□
Proof of Theorem 3.4.1

Proof. The strategy for this proof is to give an explicit transformation that separates the ELVIS moment restrictions into two subvectors: one that does not depend on the latent variables and one that has a nonsingular conditional variance matrix. In light of the results in chapters 1 and 2, efficient ELVIS is asymptotically equivalent to GMM based on the subvector that does not depend on the latent variables. Define a nonsingular linear transformation

\[
A(\theta) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & -\gamma & 1 & 0 & 0 & -\gamma \\
0 & 1 & 0 & 0 & 0 & 0 & -\gamma & 1 & 0 & -\gamma \\
0 & 0 & -\gamma I_k & 0 & 0 & 0 & 0 & I_k & 0 & 0 \\
0 & 0 & 0 & -\gamma I_k & 0 & 0 & 0 & I_k & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I_k & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]
Applying this transformation to the moment restrictions yields:

\[
A(\theta)g(U, Z, \theta) = \begin{pmatrix}
(Y - \xi'\beta - X_2\gamma)X_1 \\
(Y - \xi'\beta - X_1\gamma)X_2 \\
(Y - \xi'\beta - X_1\gamma)\xi \\
(Y - \xi'\beta - X_2\gamma)\xi \\
X_1 - X_2
\end{pmatrix}
\begin{pmatrix}
\Delta X_2 \\
\Delta X_1\Delta X_2 \\
(Y - \xi'\beta - X^*\gamma)X^* \\
(Y - \xi'\beta - X^*\gamma)\xi \\
\Delta X_1X^* \\
\Delta X_2X^*
\end{pmatrix}
= \begin{pmatrix}
g_A(Z, \theta) \\
g_B(U, Z, \theta)
\end{pmatrix}
\]

From our results we know that the efficient ELVIS estimator constructed from the moment restriction \( g(U, Z, \theta) \) will be asymptotically equivalent to the efficient GMM estimator constructed from \( g_A(Z, \theta) \) if \( E[Var(g_B(U, Z, \theta^0)|Z)] \) is nonsingular. We may further partition \( g_A(Z, \theta) \) as:

\[
g_A(Z, \theta) = \begin{pmatrix}
(Y - \xi'\beta - X_2\gamma)X_1 \\
(Y - \xi'\beta - X_1\gamma)X_2 \\
(Y - \xi'\beta - X_1\gamma)\xi \\
(Y - \xi'\beta - X_2\gamma)\xi \\
X_1 - X_2
\end{pmatrix}
\begin{pmatrix}
g^{CM}(Z, \theta) \\
X_1 - X_2
\end{pmatrix}
\]

and standard results for control variates tell us that the efficient GMM estimator constructed from \( g_1(Z, \theta) \) is asymptotically equivalent to \( g^{CM}(Z, \theta) \) if \( Cov[g^{CM}(Z, \theta^0), X_1 - X_2] = 0 \). Thus we
may conclude that the Chalfin McCrary estimator is efficient if both

\[ |E[Var(g_B(U, Z, \theta^0)|Z)]| \neq 0 \]

and \( Cov[g^{CM}(Z, \theta^0), X_1 - X_2] = 0 \).

\[ \square \]

**Proof of Theorem 3.4.2**

*Proof.* This proof is follows the same strategy as the proof of Theorem 3.4.1. For \( \gamma^0 = 0 \), the Chalfin McCrary moment restrictions are linearly dependent:

\[
g^{CM}(Z, \theta^0) = \begin{pmatrix}
(Y - \xi^\prime \beta^0)X_1 \\
(Y - \xi^\prime \beta^0)X_2 \\
(Y - \xi^\prime \beta^0)\xi \\
(Y - \xi^\prime \beta^0)\xi
\end{pmatrix}
\]

so that the Chalfin McCrary estimator is asymptotically equivalent to the GMM estimator based on:

\[
g^{CM^*}(Z, \theta^0) = \begin{pmatrix}
(Y - \xi^\prime \beta^0)X_1 \\
(Y - \xi^\prime \beta^0)X_2 \\
(Y - \xi^\prime \beta^0)\xi
\end{pmatrix}
\]
For $\gamma^0 = 0$ we use the nonsingular linear transformation

$$A^*(\theta) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & I_k & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & I_k & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & I_k & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Applying this transformation to the moment restrictions yields:

$$A^*(\theta^0)g(U, Z, \theta^0) = \begin{pmatrix}
(Y - \xi^0_\theta^0)X^1 \\
(Y - \xi^0_\theta^0)X^2 \\
\xi\Delta X^1 \\
\xi\Delta X^2 \\
X^1 - X^2 \\
\Delta X^2 \\
\Delta X^1\Delta X^2 \\
(Y - \xi^0_\theta^0)X^* \\
(Y - \xi^0_\theta^0)\xi \\
\Delta X^1 X^* \\
\Delta X^2 X^*
\end{pmatrix}$$
A permutation of previous vector is

\[
\begin{pmatrix}
(Y - \xi'\beta^0)X_1 \\
(Y - \xi'\beta^0)X_2 \\
(Y - \xi'\beta^0)\xi \\
X_1 - X_2 \\
\xi \Delta X_1 \\
\xi \Delta X_2 \\
\Delta X_2 \\
\Delta X_1 \Delta X_2 \\
(Y - \xi'\beta^0)X^* \\
\Delta X_1 X^* \\
\Delta X_2 X^*
\end{pmatrix}
\]

So similar to first case Chalfin McCrary is efficient if

\[
\det[E[Var(\begin{pmatrix}
\xi \Delta X_1 \\
\xi \Delta X_2 \\
\Delta X_2 \\
\Delta X_1 \Delta X_2 \\
(Y - \xi'\beta^0)X^* \\
\Delta X_1 X^* \\
\Delta X_2 X^*
\end{pmatrix}) | Z)] \neq 0
\]

and \( Cov[\begin{pmatrix}
(Y - \xi'\beta^0)X_1 \\
(Y - \xi'\beta^0)X_2 \\
(Y - \xi'\beta^0)\xi 
\end{pmatrix} , X_2 - X_1] = 0 \)

\[\square\]

3.7 Appendix: Figures and Tables
Figure 3.1: Monte Carlo sampling distributions for the two-step efficient ELVIS estimator, the Chalfin McCrary estimator, and the Andersson Møen estimator. Sample size $n=200$.

Table 3.1: Monte Carlo results for the two-step efficient ELVIS estimator, the Chalfin McCrary estimator, and the Andersson Møen estimator.
Figure 3.2: Monte Carlo sampling distributions for the two-step efficient ELVIS estimator, the Chalfin McCrary estimator, and the Andersson Moen estimator. Sample size $n=500$. 


Han, Sukjin and Adam McCloskey. 2017. “Estimation and inference with a (nearly) singular Jacobian.”


