Essays on Forecasting and Volatility Modelling

By

Gustavo Fruet Dias

A Thesis submitted for the degree of

Doctor of Philosophy (Ph.D.) in Economics

School of Economics and Finance

University of London

Queen Mary

June 2013

AUTHOR'S DECLARATION

I wish to declare:

No part of this doctoral dissertation, titled as "Essays on Forecasting and Volatility Modelling" and submitted to the University of London in pursuance of the degree of Doctor of Philosophy (Ph.D.) in Economics, has been presented to any University for any degree. Parts of Chapter 2 were undertaken as joint work with Prof. George Kapetanios.

Signed

Gustavo Fruet Dias

EXTENDED ABSTRACT

This thesis contributes to four distinct fields on the econometrics literature: forecasting macroeconomic variables using large datasets, volatility modelling, risk premium estimation and iterative estimators. As a research output, this thesis presents a balance of applied econometrics and econometric theory, with the latter one covering the asymptotic theory of iterative estimators under different models and mapping specifications. In Chapter 1 we introduce and motivate the estimation tools for large datasets, the volatility modelling and the use of iterative estimators.

In Chapter 2, we address the issue of forecasting macroeconomic variables using medium and large datasets, by adopting vector autoregressive moving average (VARMA) models. We overcome the estimation issue that arises with this class of models by implementing the iterative ordinary least squares (IOLS) estimator. We establish the consistency and asymptotic distribution considering the ARMA(1,1) and we argue these results can be extended to the multivariate case. Monte Carlo results show that IOLS is consistent and feasible for large systems, and outperforms the maximum likelihood (MLE) estimator when sample size is small. Our empirical application shows that VARMA models outperform the AR(1) (autoregressive of order one model) and vector autoregressive (VAR) models, considering different model dimensions.

Chapter 3 proposes a new robust estimator for GARCH-type models: the nonlinear iterative least squares (NL-ILS). This estimator is especially useful on specifications where errors have some degree of dependence over time or when the conditional variance is misspecified. We illustrate the NL-ILS estimator by providing algorithms that consider the GARCH(1,1),

weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. I establish the consistency and asymptotic distribution of the NL-ILS estimator, in the case of the GARCH(1,1) model under assumptions that are compatible with the quasi-maximum likelihood (QMLE) estimator. The consistency result is extended to the weak-GARCH(1,1) model and a further extension of the asymptotic results to the GARCH(1,1)-in-mean case is also discussed. A Monte Carlo study provides evidences that the NL-ILS estimator is consistent and outperforms the MLE benchmark in a variety of specifications. Moreover, when the conditional variance is misspecified, the MLE estimator delivers biased estimates of the parameters in the mean equation, whereas the NL-ILS estimator does not. The empirical application investigates the risk premium on the CRSP, S&P500 and S&P100 indices. I document the risk premium parameter to be significant only for the CRSP index when using the robust NL-ILS estimator. We argue that this comes from the wider composition of the CRPS index, resembling the market more accurately, when compared to the S&P500 and S&P100 indices. This finding holds on daily, weekly and monthly frequencies and it is corroborated by a series of robustness checks.

Chapter 4 assesses the evolution of the risk premium parameter over time. To this purpose, we introduce a new class of volatility-in-mean model, the time-varying GARCH-in-mean (TVGARCH-in-mean) model, that allows the risk premium parameter to evolve stochastically as a random walk process. We show that the kernel based NL-ILS estimator successfully estimates the time-varying risk premium parameter, presenting a good finite sample performance. Regarding the empirical study, we find evidences that the risk premium parameter is time-varying, oscillating over negative and

positive values.

Chapter 5 concludes pointing the relevance of of the use of iterative estimators rather than the standard MLE framework, as well as the contributions to the applied econometrics, financial econometrics and econometric theory literatures.

ACKNOWLEDGEMENT

I would like to thank my supervisor Professor George Kapetanios for his guidance. He believed in my potential five years ago, and therefore I am grateful to him.

I would like to thank my family. In special, my parents Maria Isabel Fruet Dias and João Serratti Dias, who have prioritized my education and, most important of all, have always pushed me to become a better person. I owe them this and several other achievements in my life. They inspired me to always work hard. I will be forever indebted to them.

Finally, I want to thank my beloved wife and best friend Cristina Mabel Scherrer. I would definitely not be writing this thesis and finishing the Ph.D. programme without her support, understanding, lovely smile and friendship. Her strength and determination have always inspired me through this journey. I will be forever indebted to her.

"You will never do anything in this world without courage. It is the greatest quality of the mind next to honor." **Aristotle**

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis investigates the use of iterative estimators, with applications to two distinct fields: the applied econometrics and the financial econometrics fields. We therefore present a balance between econometric theory and empirical results covering topics such as forecasting and volatility modelling. Our contribution to the econometric theory literature consists in establishing the asymptotic theory for two variants of iterative estimators (the iterative ordinary least squares estimator (IOLS) and the nonlinear iterative least squares estimator (NL-ILS)). We derive theoretical results for two of the most important time series models adopted in the literature: the autoregressive moving average (ARMA) and the generalized autoregressive conditional heteroscedasticity (GARCH) models. The two alternative iterative estimators we adopt in this thesis overcome estimation issues related with the vector autoregressive moving average (VARMA), GARCH, weak-GARCH, and GARCH-in-mean models. In general lines, our empirical application sheds light on the validity of VARMA models on forecasting key macroeconomic variables using large datasets (Chapter 2), as well as

on the identification of the risk-return tradeoff (Chapters 3 and 4).

Forecasting macroeconomic variables received great attention in the economic literature in the past decades. Time series econometrics played a major role in this process, following the seminal paper of Sims (1980) and the wide implementation of vector autoregressive (VAR) models. A more recent extension of this literature relates to forecasting key macroeconomic variables using large datasets. These large datasets became more widely available in the past years and their use is motivated by the intuition they should reflect agents's information set more appropriately. Hence, by incorporating large datasets into econometric models, forecast accuracy should improve. The challenge of dealing with large datasets comes because standard econometric frameworks usually lose performance when the number of variables (parameters) increases, the so-called "curse of dimensionality".

Potential solutions for this problem arise from mainly two different group of models: penalized regressions and factor models. The first group, penalized regressions, aims to overcome the dimensionality issue by imposing restrictions on the parameter matrices of a standard VAR model. The intuition behind this solution arises from a well-known result of standard linear regression which states that covariance matrices of restricted estimators have lower variances than those of unrestricted estimators. Among the many important contributions from this field, we point out the following classes of models: Bayesian VAR (BVAR) (De Mol, Giannone, and Reichlin (2006) and Banbura, Giannone, and Reichlin (2007)), in the spirit of Doan, Litterman, and Sims (1984) and Litterman (1986); Ridge (De Mol, Giannone, and Reichlin (2006)) and shrinkage estimators (Carriero, Kapetanios, and Marcellino (2008)); Reduced Rank VAR (Carriero, Kapetanios, and

Marcellino (2011)); and Lasso (De Mol, Giannone, and Reichlin (2006) and Tibshirani (1996)).

The second group of models dealing with the "curse of dimensionality"is the factor models. The seminal works in this area are Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002). Factor models summarize a large number of variables with only a few unobserved common factors and an idiosyncratic component. These models dramatically reduce the dimensions of the system, contributing to an improvement in forecast accuracy. Common factor models improve forecast accuracy and produce theoretically well-behaved impulse response functions, as reported by De Mol, Giannone, and Reichlin (2006) and Bernanke, Boivin, and Eliasz (2005). These findings support the idea that agents consider wider information sets when making their decisions.

Alternatively to the methodologies discussed above, we propose the use of vector autoregressive moving average (VARMA) models to address the "curse of dimensionality". The intuition supporting this choice is that VARMA models share features from both penalized regressions and factor models. The first is the reduction of the model dimensionality, achieved by setting some elements of the parameter matrices to zero following uniqueness requirements. The second is the parsimonious summarizing of high-order autoregressive lags into low-order lagged shocks. By adopting VARMA, we allow lagged shocks from most of the macroeconomic variables in our dataset to play very important roles in forecasting the future realizations of key macroeconomic variables. We overcome the estimation issue that arises with VARMA models by implementing the IOLS estimator. We establish the consistency and asymptotic distribution considering the

ARMA(1,1) and we argue these results can be extended to the multivariate case. Monte Carlo results show that IOLS is consistent and feasible for large systems, and outperforms the maximum likelihood (MLE) estimator when sample size is small. Our empirical application shows that VARMA models outperforms the AR(1) and VAR models, considering different model dimensions.

The second part of this thesis deals with volatility modelling and risk premium estimation. Time-varying volatility plays a major role in both finance and economics. In particular, asset return volatility is paramount in fields such as asset pricing, risk management and portfolio allocation. The task of modeling the conditional variance has been a central topic in econometrics following the seminal papers of Engle (1982) and Bollerslev (1986). Since then, different specifications and frameworks, such as GARCH-type models, stochastic volatility, realized volatility and combinations of these approaches have been adopted, trying to capture the very specific stylized facts observed in financial returns. A natural extension that emerges from modeling the conditional variance is the relation between risk and return. The intertemporal capital asset pricing model (ICAPM) of Merton (1973) establishes a positive relation between the conditional excess returns and the conditional variance, implying that investors should be remunerated for bearing extra risk. In spite of its simple specification, empirical evidences on the sign and significance of the risk premium parameter are blurred. Bollerslev, Chou, and Kroner (1992), Lettau and Ludvigson (2010), Rossi and Timmermann (2010), among others highlight three potential problems that contribute to the lack of consensus regarding the existence of the risk-return tradeoff.

First, quasi-maximum likelihood (QMLE) estimates of the risk premium parameter using the GARCH-in-mean framework may be inconsistent if the conditional variance is misspecified. Hence, given the vast menu of alternative volatility models available in the literature, it is paramount to use estimators which are robust to a large number of volatility specifications. Secondly, misspecification of the risk premium function may lead to biased results. Third, the use of only few conditioning variables generates incomplete models, making very difficult the identification of the risk premium function. Chapter 3 addresses the first issue raised above, whereas Chapter 4 deals explicitly with the second and third issues.

1.1 Outline of Thesis

Chapter 2 addresses the issue of forecasting key macroeconomic variables using medium and large datasets (from 10 to 40 variables). As an alternative to standard autoregressive (AR) and vector autoregressive (VAR) models, we propose using VARMA models. We overcome the estimation issue that usually arises in high dimensional VARMA models by adopting the IOLS estimation procedure. We establish consistency and the asymptotic distribution for the IOLS estimator considering the ARMA(1,1) case, providing an analytical expression for the latter one. We report results from Monte Carlo simulations, assessing the consistency, efficiency, and forecast accuracy obtained using the IOLS estimator. With regard to the consistency and efficiency analysis, we show the IOLS estimator is consistent and feasible for large systems, and also performs better than the maximum-likelihood estimator (MLE) when sample size is small. In terms

of forecast accuracy, we report an outstanding performance of VARMA models compared with VAR and AR(1) models under a variety of specifications. On the empirical application, we show that different specifications of VARMA models estimated using the IOLS framework provide more accurate forecasts than VAR and AR(1) models, considering different model dimensions.

Chapter 3 investigates the significance of the risk premium parameter in three different market indices. We propose a novel full parametric iterative estimator, the NL-ILS estimator, nesting several GARCH-type models. This estimator is especially useful on specifications where errors have some degree of dependence over time or the conditional variance is misspecified. We derive the asymptotic theory for the GARCH(1,1) and weak-GARCH(1,1) models under assumptions that are compatible with the QMLE estimator. We argue that these results can be extended to different GARCH-type models. A Monte Carlo study provides evidences that the NL-ILS estimator is consistent and outperforms the MLE benchmark in a variety of specifications. Moreover, when the conditional variance is misspecified, the MLE estimator delivers biased estimates of the parameters in the mean equation, whereas the NL-ILS estimator does not. We report an outstanding performance of the NL-ILS estimator when estimating volatility models generated with time dependent innovations.

We examine the significance of the risk premium parameter using the GARCH(1,1)-in-mean framework by adopting the NL-ILS estimator. The main question is whether, by using an estimator which is robust to misspecification of the conditional variance, the risk premium parameter is significant and presents the correct sign. We assess this question in two

different dimensions: temporal frequency and market *proxy*. The former one is evaluated by estimating the model on a daily, weekly and monthly basis, whereas the latter dimension is appraised by adopting three different indices: CRSP, S&P500 and S&P100. The choice of comparing different indices emerges from the distinctive compositions they have. The CRSP index is known to be the best *proxy* for the market, whereas S&P100 would be the least complete index. By estimating the risk premium at different frequencies, we control for the QMLE lack of consistency that arises when the considered sampling frequency is different from the true data generation process. We find significant risk premium parameter only for the CRSP index when using the robust NL-ILS estimator. We obtain a different picture with QMLE: the risk premium parameter is significant for all indices, including the least complete one, the S&P100. The significance of the risk premium parameter when estimated with the NL-ILS holds on daily, weekly and monthly frequencies and it is corroborated by a series of robustness checks. We argue that the NL-ILS estimator is the only one able to capture the "true" risk premium, since its results reflect the wider composition of the CRPS index, resembling the market more accurately, when compared to S&P500 and S&P100 indices.

Chapter 4 examines how the risk premium parameter varies over time, shedding light on the behaviour of the risk aversion parameter during periods of financial distress. To accommodate a time-varying coefficient on the mean equation of a GARCH-in-mean model, we introduce the time-varying GARCH-in-mean (TVGARCH-in-mean) model, where the risk premium parameter is allowed to be a time-varying stochastic process. We propose an estimation strategy that combines kernel methods with the NL-ILS esti-

mator and successfully estimates the time-varying risk premium parameter. A Monte Carlo study shows that the proposed algorithm has good finite sample properties. We investigate the time-varying risk premium using excess returns on the CRSP index. We document that the risk premium parameter is indeed time-variant and shows high degree of persistence. We find that the monthly time-varying risk premium parameter is statistically different from zero on 46.5% of the observations. Considering point-wise analyses, we find that weekly estimates of the time-varying risk premium parameter anticipate *bear market* phases and business cycles fluctuations. Finally, our results suggest that the relation between significance of the time-varying risk premium parameter and business cycle fluctuations has changed in the past twenty years.

Chapter 5 draws the conclusion and final remarks of the thesis.

Chapter 2

Forecasting Medium and Large Datasets with Vector Autoregressive Moving Average (VARMA) Models

2.1 Introduction

The use of large arrays of economic indicators to forecast key macroeconomic variables has become very popular recently. Economic agents consider a wide range of information when they construct their expectations about the behavior of macroeconomic variables such as interest rates, industrial production, and inflation. In the past several years, this information has become more widely available through a large number of indicators that aim to describe different sectors and fundamentals from the whole economy. To improve forecast accuracy, large datasets that attempt to replicate the

set of information used by agents to make their decisions are incorporated into econometric models.

For the past twenty years, macroeconomic variables have been forecasted using vector autoregression (VAR) models. This type of models performs well when the number of variables in the system is relatively small. When the number of variables increases, however, the performance of VAR forecasts deteriorates very fast, generating the so-called "curse of dimensionality". The reasons for this problem are: first, some variables in the VAR models are not Granger-Caused by some components of the system; and second, the sample data is not rich enough. In both cases, large errors are associated with the parameter estimates, contributing to the reduced forecast accuracy of this class of models.

In this chapter, we propose the use of vector autoregressive moving average (VARMA) models, estimated using iterative ordinary least squares (IOLS) estimator, as a feasible method to address the "curse of dimensionality"on medium and large datasets. VARMA models have been studied for the past thirty years, but they have not been, by far, as popular as VAR models. The most recent attempt to use the VARMA approach to forecast macroeconomic variables comes from Athanasopoulos and Vahid (2008). They applied the VARMA methodology to small systems (three- and four-variable models) and obtained forecasts that were better than those obtained using standard VAR models. As far as our knowledge goes, the VARMA methodology has never been applied to medium and large datasets, as we do in this chapter. There are two main issues that contribute to the scarcity of VARMA models in the literature: estimation and specification. In this chapter, we tackle the first issue, by proposing the

use of IOLS in the spirit of Kapetanios (2003). We show, through Monte Carlo simulations, that the standard estimation procedure for VARMA models (maximum-likelihood estimator (MLE)) is not feasible for systems with more than eight variables, whereas the IOLS estimator is feasible and consistent even for high-dimensional models.

Other methodologies have been proposed in the literature to deal with the "curse of dimensionality". The first group of models, penalized regressions, aims to overcome the dimensionality issue by imposing restrictions on the parameter matrices of a standard VAR model. The intuition behind this solution arises from a well-known result from standard linear regression which states that covariance matrices of restricted estimators have lower variances than those of unrestricted estimators. Among the many important contributions from this field, we point out the following classes of models: Bayesian VAR (BVAR) (De Mol, Giannone, and Reichlin (2006) and Banbura, Giannone, and Reichlin (2007)), in the spirit of Doan, Litterman, and Sims (1984) and Litterman (1986); Ridge (De Mol, Giannone, and Reichlin (2006)) and shrinkage estimators (Carriero, Kapetanios, and Marcellino (2008)); Reduced Rank VAR (Carriero, Kapetanios, and Marcellino (2011)); and Lasso (De Mol, Giannone, and Reichlin (2006) and Tibshirani (1996)).

The second group of models dealing with the "curse of dimensionality"is the factor models. The seminal works in this area are Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002). Factor models summarize a large number of variables with only a few unobserved common factors and an idiosyncratic component. These models dramatically reduce the dimensions of the system, contributing to an improvement in

forecast accuracy. Common factor models improve forecast accuracy and produce theoretically well-behaved impulse response functions, as reported by De Mol, Giannone, and Reichlin (2006) and Bernanke, Boivin, and Eliasz (2005). These findings support the idea that agents consider wide sets of information when making their decisions.

VARMA models are able to capture two important features from the penalized regressions and the common factor models. The first is the reduction of the model dimensionality, achieved by setting some elements of the parameter matrices to zero following uniqueness requirements. The second is the parsimonious summarizing of high-order autoregressive lags into low-order lagged shocks. By adopting VARMA, we allow lagged shocks from most of the macroeconomic variables in our dataset to play very important roles in forecasting the future realizations of key macroeconomic variables.

With regard to the theory, we establish the consistency and asymptotic distribution of the IOLS estimator by considering the univariate ARMA(1,1) model. Our asymptotic results are obtained under mild assumptions using the asymptotic contraction mapping framework defined in Dominitz and Sherman (2005). We argue that these theoretical results can be extended to VARMA models. To support this claim, we provide an extensive Monte Carlo study showing that IOLS estimator is consistent under different system dimensions and specifications. Furthermore, we show that, compared to the MLE estimator, the IOLS procedure delivers outstanding gains in terms of mean squared error when the sample size is small.

In our empirical application, we report results from three different system sizes: 10, 20, and 40 variables. We design five different datasets taken from Stock and Watson (2005) for each system dimension. We evaluate

the first and fourth out-of-the-sample forecast performances of VARMA models, comparing them with standard VAR(1) and AR(1) models; the latter is considered one of the benchmark models for the Stock and Watson (2005) dataset. The VARMA framework produces competitive forecasts, especially for longer horizons. We show that VARMA models produce more accurate forecasts than the AR(1) benchmark does, considering different system sizes and specifications. In particular, we point out that VARMA models compare favorably with their competitors when the dataset is large (40 variables).

The chapter is structured as follows. In Section 2.2, we discuss the properties of VARMA models and derive the IOLS estimator. In Section 2.3, we establish the consistency and asymptotic distribution of the IOLS estimator. In Section 2.4, we address the consistency, efficiency, and forecast accuracy of VARMA models estimated with the IOLS procedure through a Monte Carlo study. In Section 2.5, we display the results from our empirical application. The Appendix displays the proofs.

2.2 VARMA Models and Estimation Procedures

Our interest lies in forecasting key elements of the K dimensional vector process $Y_t = (y_{1,t}, y_{2,t}, ..., y_{K,t})'$, where K is allowed to be large. We assume, as a baseline model, a general VARMA(p,q) model where the means have been removed. The disturbances $u_t = (u_{1,t}, u_{2,t}, ..., u_{K,t})'$ are assumed to be a zero-mean white-noise process with a non-singular covariance matrix

$u_t \sim (0, \Sigma_u)$.

$$A_0 Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + M_0 u_t + M_1 u_{t-1} + ... + M_q u_{t-q} \quad (2.1)$$

The baseline model stated in (2.1) can be rewritten in two different forms: lag notation form (2.2) and compact form (2.3). These representations will be very useful for deriving some important theoretical properties, as well as for the estimation procedure.

$$A(L) Y_t = M(L) u_t \quad (2.2)$$

$$Y = BX + U \quad (2.3)$$

The lag polynomials in (2.2) have the standard form: $A(L) = A_0 - A_1 L - A_2 L^2 - ... - A_p L^p$ and $M(L) = M_0 + M_1 L + M_2 L^2 + ... + M_q L^q$, where L is the lag operator. From (2.3), Y has dimension $(K \times T)$; $B = [(I_K - A_0), A_1, ..., A_p, (M_0 - I_K), M_1, ..., M_q]$ joints the parameter matrices with dimension $(K \times K(p+q+2))$; $X = (X_0, ..., X_T)$ can be seen as the matrix of regressors with dimension $(K(p+q+2) \times T)$, where $X_t = [Y_t, Y_{t-1}, ..., Y_{t-p}, U_t, U_{t-1}, ..., U_{t-q}]'$; and U is a $(K \times T)$ matrix of disturbances.

Our baseline model is assumed to be stable and invertible, and the latter is crucial in our estimation process. A general VARMA(p,q) is considered stable and invertible if $\det(A_0 - A_1 z - A_2 z^2 - ... - A_p z^p) \neq 0 \; for \; |z| \leq 1$ and $\det(M_0 - M_1 z - M_2 z^2 - ... \; - M_p z^q) \neq 0 \; for \; |z| \leq 1$ hold, respectively. If the model is invertible, it is possible to express the VARMA(p,q)

as an infinite standard VAR process as follows:

$$\Pi_0 Y_t = \sum_{i=1}^{\infty} \Pi_i Y_{t-i} + u_t \tag{2.4}$$

where $\Pi(L) = M(L)^{-1} A(L)$.

The result in (2.4) is extremely important in two ways. On one hand, it will be an important tool in deriving our estimation procedure and consistency proofs. On the other hand, it gives the intuitive reason why a VARMA model outperforms a VAR specification when forecasting large datasets: an invertible VARMA(p,q) with finite p and q can be a parsimonious representation of a very long (infinite) VAR process. In other words, if the true data generation process is a VARMA(p,q) process, then fitting a VAR(p) would lead to the estimation of pK^2 parameters. Considering a large K, as is done in this study, it would cause forecast accuracy to deteriorate very fast. In contrast to the VAR case, VARMA models require some particular conditions to assure that the model is unique. There are different transformations that guarantee uniqueness for the VARMA class of models. Athanasopoulos, Poskitt, and Vahid (2007) show that VARMA models specified using scalar components perform slightly better in empirical exercises than ones using the Echelon Form methodology. The authors claim, however, that the latter has the advantage of having a simpler identification procedure. In this chapter, we implement the Echelon Form transformation as a way to impose uniqueness in both Monte Carlo and empirical applications.

A general VARMA model such as the one stated in (2.1) is considered to be in its Echelon Form when there are no common factors on the polyno-

mials $A(L)$ and $M(L)$ and the conditions stated in equations (2.5), (2.6), (2.7), (2.8) and (2.9) are satisfied (see Lütkepohl (2007) pg. 452 for more details).

$$p_{ki} = \begin{cases} \min(p_k + 1, p_i) & \text{for } k \geq i \\ \max(p_k, p_i) & \text{for } k < i \end{cases} \tag{2.5}$$

$$\alpha_{kk}(L) = 1 - \sum_{j=1}^{p_k} \alpha_{kk,j} L^j, \qquad \text{for } k = 1, ..., K \tag{2.6}$$

$$\alpha_{ki}(L) = - \sum_{j=p_k - p_{ki}+1}^{p_k} \alpha_{ki,j} L^j \qquad \text{for } k \neq i \tag{2.7}$$

$$m_{ki}(L) = -\sum_{j=0}^{p_k} m_{ki,j} L^j, \qquad \text{for } k = 1, ..., K \tag{2.8}$$

$$M_0 = A_0 \tag{2.9}$$

where $A(L)=[\alpha_{ki}]_{k,i=1,...,K}$ and $M(L)=[m_{ki}]_{k,i=1,...,K}$ are, respectively, the operators from the autoregressive and moving average components of the VARMA process. The arguments $[p_k]_{k,i=1,...,K}$ are Kronecker Indices and denote the maximum degrees of both polynomials $A(L)$ and $M(L)$, being exogenously defined. The p_{ki} numbers can be interpreted as the free coefficients in each operator $\alpha_{ki}(L)$ for $i \neq k$ from the $A(L)$ polynomial. By imposing restrictions on the coefficient matrices due to the Echelon Form transformation, VARMA models have the desirable feature that many of the coefficients from both the autoregressive and moving average matrices are equal to zero.

VARMA models, similar to their univariate (ARMA model) counterparts, are usually estimated using the MLE procedure. Provided that the model in (2.1) is uniquely defined and disturbances U_t are normally dis-

tributed, MLE delivers consistent and efficient estimators. Although MLE seems to be very powerful at first glance, it presents serious problems when dealing with VARMA models that account for medium and large datasets. We report the results of Monte Carlo simulations in Section 2.4 that demonstrate how MLE becomes hardly feasible for VARMA models with more than eight variables. We overcome this issue by implementing an IOLS procedure in the spirit of Kapetanios (2003), who shows that IOLS estimators compare favorably with MLE estimators for ARMA models and a bivariate VARMA(1,1) model. In this chapter we go much further in three different directions: first, by establishing the asymptotic theory for the IOLS estimator under assumptions compatible with the quasi-maximum-likelihood (QMLE) estimator; second, by showing, through an extensive Monte Carlo, that the theory developed for the univariate case can be extended to high-dimensional VARMA models; third, by assessing forecast performance of VARMA models, estimated with IOLS, compared with autoregressive (AR) and VAR models under different system dimensions.

The IOLS framework consists of computing ordinary least squares (OLS) estimates of the parameters using estimates of the latent regressors. These regressors are computed recursively at each iteration using the OLS estimates as functions of the parameters. Under the VARMA setup, we are interested in estimating the parameter matrices $A_0, A_1, ..., A_p, M_0, ...,$ and M_q.

Following the uniqueness discussion, we assume that the model in (2.1) is expressed in its Echelon Form and is therefore uniquely defined. Echelon Form transformation implies that $A_0 = M_0$, which leads to a different specification of matrices in (2.10) when compared with the compact notation

displayed in (2.3).

$$vec\,(Y) = (X' \otimes I_K)\, vec\,(B) + vec\,(U) \tag{2.10}$$

We now have that $B = [(I_K - A_0), A_1, ..., A_p, , M_1, ..., M_q]$ with dimension $(K \times K(p+q+1))$; $X = (X_0, ..., X_T)$ is the matrix of regressors with dimension $(K(p+q+1) \times T)$, where $X_t = [Y_t - U_t, Y_{t-1}, ..., Y_{t-p}, U_{t-1}, ..., U_{t-q}]'$; and U is a $(K \times T)$ matrix of disturbances. Note that the matrices of parameters may not be full matrices because Echelon Form transformation can set many of their elements to zero.

Rewriting the matrix of parameters $vec\,(B)$ into the product of a matrix R, that accounts for the restrictions from the Echelon Form transformation, and a vector β, that joints the free parameters, shows that β could easily be estimated with OLS in the case that regressors were fully observed.

$$vec\,(Y) = (X' \otimes I_K)\, R\beta + vec\,(U) \tag{2.11}$$

The matrix X in (2.11), however, is not fully observed; it contains lagged values of the latent disturbances. Using the invertibility condition, we can express a finite VARMA model into an infinite VAR as stated in (2.4). The only difference from (2.4) arises from the Echelon Form transformation, which imposes $\Pi_0 = I_K$.

We compute estimates of U by truncating (2.4) into some lag order p that minimizes the AIC criterion, as in (2.12). Following the result from Ng and Perron. (1995), this procedure delivers consistent estimates of U,

and we denote them as $\widehat{U}^0$.

$$\widehat{U}_t^0 = Y_t - \sum_{i=1}^{p} \widehat{\Pi}_i Y_{t-i} \tag{2.12}$$

By substituting $\widehat{U}^0$ into the matrix X in (2.11), we denote this matrix of regressors as $\widehat{X}^0$ because it contains the first estimates of the lagged latent disturbances. The first iteration in our IOLS method is obtained by computing the OLS estimator from the modified version of (2.11). The vector of parameter estimates $\widehat{\beta}^1$ in (2.13) is therefore the first iteration from the IOLS algorithm.

$$\widehat{\beta}^1 = \left[R' \left(\widehat{X}^0 \widehat{X}^{0\prime} \otimes I_K\right) R\right]^{-1} R' \left(\widehat{X}^{0\prime} \otimes I_K\right) vec(Y) \tag{2.13}$$

We are now in a position to use $\widehat{\beta}^1$ to recover the parameter matrices $\widehat{A}_0^1, ..., \widehat{A}_p^1$, $\widehat{M}_1^1, ..., \widehat{M}_q^1$ and a new set of residuals $\widehat{U}^1$ by recursively applying (2.14). Note that the superscript on the parameter matrices refers to the iteration in which those parameters were computed, whereas the subscript is the usual lag order.

$$\begin{aligned} \widehat{U}_t^1 = \left[\widehat{A}_0^1\right]^{-1} \Big[\widehat{A}_0^1 Y_t - \widehat{A}_1^1 Y_{t-1} - ... \widehat{A}_p^1 Y_{t-p} - \widehat{M}_1^1 \widehat{U}_{t-1}^1 - \\ ... - \widehat{M}_q^1 \widehat{U}_{t-q}^1\Big] \end{aligned} \tag{2.14}$$

We compute the second iteration of the IOLS procedure by plugging $\widehat{U}_t^1$ into (2.11) yielding $\widehat{X}^1$. Note that $\widehat{X}^1 = (\widehat{X}_0^1, ..., \widehat{X}_T^1)$, where $\widehat{X}_t^1 = [Y_t - \widehat{U}_t^1, Y_{t-1}, ..., Y_{t-p}, \widehat{U}_{t-1}^1, ..., \widehat{U}_{t-q}^1]'$, is a function of the estimates obtained in the first iteration: $\widehat{\beta}^1$. Using (2.13), we obtain $\widehat{\beta}^2$ and its correspondent set of residuals recursively through (2.14). The j^{th} iteration of the IOLS

estimator is thus given by (2.15), whereas its correspondent recursive residuals are given by (2.16).

$$\begin{aligned}\widehat{\beta}^{j(T)} &= \widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right) = \\ &= \left[R'\left(\widehat{X}^{j-1}\widehat{X}^{j-1\prime}\otimes I_K\right)R\right]^{-1} R'\left(\widehat{X}^{j-1\prime}\otimes I_K\right) vec(Y)\end{aligned} \tag{2.15}$$

$$\widehat{U}_t^j = \left[\widehat{A}_0^j\right]^{-1}\left[\widehat{A}_0^j Y_t - \widehat{A}_1^j Y_{t-1} - ...\widehat{A}_p^j Y_{t-p} - \widehat{M}_1^j\widehat{U}_{t-1}^j - ... - \widehat{M}_q^j\widehat{U}_{t-q}^j\right] \tag{2.16}$$

We stop the IOLS algorithm when estimates of β converge. In both the empirical application and the Monte Carlo study, we assume that $\widehat{\beta}^j$ converges if $\| \widehat{U}_t^j - \widehat{U}_t^{j-1} \| \leq 10^{-5}$. In accordance with the notation that we adopt in Section 2.3, we will allow the number of iterations to be a function of the sample size. To simplify our notation on $\widehat{U}_t^j$ and $\widehat{X}^j$, we only make it explicit that $j(T)$ is a function of T when we denote estimates of β that were obtained from the IOLS algorithm. The rate at which $j(T)$ needs to increase as $T \longrightarrow \infty$ will be discussed further in Section 2.3. We denote the function $\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right)$ as the sample mapping for the IOLS estimator. The sample mapping $\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right)$ maps $\widehat{\beta}^{j(T)-1}$ to $\widehat{\beta}^{j(T)}$.

In one particular case, the IOLS estimator algorithm does not converge. This arises when some iteration of the algorithm generates a non-invertible model. In other words, as discussed in Kapetanios (2003), the algorithm will not converge if the mapping stated in (2.15) is not a contraction mapping. A similar discussion arises in Dominitz and Sherman (2005). They prove that a compulsory condition for convergence comes from guaranteeing that the mapping $\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right)$ is an asymptotic contraction mapping

(ACM)[1]. As pointed out by Dominitz and Sherman (2005), if a collection is an ACM, then it will have an unique fixed point in $(\mathbb{B}, d)$, where the fixed point now depends on the sample characteristics, i.e., of each T and ω, with $\omega \in \Omega$ and Ω being the sample space. Therefore, non-convergence of the IOLS algorithm in a finite sample may be caused by a small T, a sample that yields a mapping that is not an ACM, or by some particular combinations of the eigenvalues governing the autoregressive and moving average parameter matrices on the true data generation process (DGP). It is important to point out that even in such cases, $\widehat{\beta}^1$ is a consistent estimate of β, following the fact that $\widehat{U}_t^0$ converges to U_t for all t.

2.3 Theoretical Properties

This section provides theoretical results regarding the consistency and asymptotic distribution of the IOLS estimator discussed in the previous sections. As a matter of simplicity, we focus our analysis on the univariate ARMA class of models. Extension to the VARMA case will be discussed further in this section. We base our results on Dominitz and Sherman (2005). We define an ARMA(1,1) model as

$$y_t = \beta_1 y_{t-1} + u_t + \beta_2 u_{t-1} \tag{2.17}$$

$$y_t = X_{-1,t}\beta + u_t \tag{2.18}$$

$$Y = X_{-1}\beta + U \tag{2.19}$$

[1]From their definition, a collection $\{K_T^{\omega}(.) : T \geq 1, \omega \in \Omega\}$ is an ACM on $(\mathbb{B}, d)$ if $d\left(K_T^{\omega}(x), K_T^{\omega}(y)\right) \leq cd(x, y)$ as $T \longrightarrow \infty$, where $c \in [0, 1)$, $(\mathbb{B}, d)$ is a metric space with $x, y \in \mathbb{B}$, $(\Omega, \mathcal{A}, \mathcal{P})$ denoting a probability space and $K_T^{\omega}(.)$ is a function defined on $\mathbb{B}$.

Equation (2.18) expresses the model in (2.17) in a compact notation, where $X_{-1,t} = [y_{t-1}, u_{t-1}]$, whereas (2.19) adopts the standard matrix notation, with both $Y = [y_1, y_2, ..., y_T]'$ and $U = [u_1, u_2, ..., u_T]'$ being $(T \times 1)$ vectors, $X_{-1} = [Y_{-1}, U_{-1}]$ a $(T \times 2)$ matrix and $\beta = (\beta_1, \beta_2)'$.

To derive consistency and asymptotic distribution of the IOLS estimator for a ARMA(1,1) model stated in (2.17), we impose the following assumptions:

Assumption 1 (Stability, Invertibility) *The model in (2.17) is stable, invertible, and contains no common factors, i.e.,* $|\beta_1| < 1$, $|\beta_2| < 1$ *and* $\beta_1 \neq -\beta_2$.

Assumption 2 (Disturbances) *The disturbance* u_t *in (2.17) is independent and identically distributed (iid) process with* $\mathbb{E}(u_t) = 0$, $Var(u_t) = \sigma_u^2$ *and finite fourth moment.*

Definition 1 (Mapping) *We define the sample mapping* $\widehat{N}_T\left(\widehat{\beta}^{j(T)}\right)$ *and its population counterpart* $N\left(\beta^{j(T)}\right)$ *as follows:*

i. $\widehat{\beta}^{j(T)+1} = \widehat{N}_T\left(\widehat{\beta}^{j(T)}\right) = \left[\frac{\widehat{X}_{-1}^{j\prime}\widehat{X}_{-1}^{j}}{T}\right]^{-1}\left[\frac{\widehat{X}_{-1}^{j\prime}Y}{T}\right]$

ii. $\beta^{j(T)+1} = N\left(\beta^{j(T)}\right) = \mathbb{E}\left[\frac{X_{-1}^{j\prime}X_{-1}^{j}}{T}\right]^{-1}\mathbb{E}\left[\frac{X_{-1}^{j\prime}Y}{T}\right]$

where $\widehat{X}_{-1}^{j}$ *and* X_{-1}^{j} *denote that regressors computed on the* j^{th} *iteration are functions of* $\widehat{\beta}^{j(T)}$ *and* $\beta^{j(T)}$*, respectively.*

$\widehat{N}_T\left(\widehat{\beta}^{j(T)}\right)$ maps from $\mathbb{R}^2$ to $\mathbb{R}^2$, and the superscript j denotes the iteration which parameters where computed. We allow j to be a function of T, in such a way that $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$. The vector $\widehat{\beta}^{j(T)} = \left(\widehat{\beta}_1^{j(T)}, \widehat{\beta}_2^{j(T)}\right)'$ joints estimates of β_1 and β_2 obtained in the j^{th}

iteration from the sample mapping in Definition 1. Note that $\widehat{\beta}^{j(T)+1}$ is the solution obtained from the minimization of the sample objective function $\widehat{Q}_T\left(\widehat{\beta}^{j(T)+1}\right)$ subject to $\widehat{\beta}^{j(T)+1} \in \mathbb{B}$, where $\mathbb{B}$ is the set of all possible parameter values satisfying Assumption 1.

$$\widehat{Q}_T\left(\widehat{\beta}^{j(T)+1}\right) = T^{-1}\sum_{t=1}^{T}\left[y_t - \widehat{X}^j_{-1,t}\widehat{\beta}^{j(T)+1}\right]^2 \tag{2.20}$$

The population mapping is the closed solution from the minimization of the population counterpart of $\widehat{Q}_T\left(\widehat{\beta}^{j(T)+1}\right)$ defined as:

$$Q\left(\beta^{j(T)+1}\right) = \mathbb{E}\left(T^{-1}\sum_{t=1}^{T}\left[y_t - X^j_{-1,t}\beta^{j(T)+1}\right]^2\right) \tag{2.21}$$

Note that as an identification condition, we have $N(\beta) = \beta$, which implies that when evaluated on the true vector of parameters, the population mapping maps the vector β to itself. This implies that if the population mapping is an ACM then β is a fixed point of $N(\beta)$. The dependency between X^j_{-1} and $\beta^{j(T)}$ arises from the fact that the j^{th} estimates of the unobserved disturbances, U^j_{-1}, are obtained using the estimates $\beta^{j(T)}$, as in (2.22). To highlight this dependence, we denote regressors and residuals, which are functions of $\beta^{j(T)}$, as X^j_{-1} and U^j_{-1}, respectively; whereas $\widehat{X}^j_{-1}$ and $\widehat{U}^j_{-1}$ are the quantities computed using $\widehat{\beta}^{j(T)}$.

$$U^j_{-1} = \left(1 + \beta_2^{j(T)}L\right)^{-1}\left(1 - \beta_1^{j(T)}L\right)Y_{-1} \tag{2.22}$$

$$\widehat{U}^j_{-1} = \left(1 + \widehat{\beta}_2^{j(T)}L\right)^{-1}\left(1 - \widehat{\beta}_1^{j(T)}L\right)Y_{-1} \tag{2.23}$$

Note that with Assumption 2 and the definition of our baseline model, we satisfy the assumptions of the OLS estimator when the regression ac-

commodates stochastic regressors such as lagged values of the dependent variable. This allows us to use OLS at each j^{th} iteration in the IOLS algorithm. From Assumption 2, note that we do not have to impose any particular distribution on the disturbances. We only require u_t to have a finite fourth moment and a continuous distribution. This is an important advantage of our setup compared with the standard MLE approach. Moreover, both the consistency and asymptotic distribution results also hold when we weaken Assumption 2, such as where u_t is a weak white noise process. If u_t is set to be a linear projection, yielding a weak ARMA model as discussed in Drost and Nijman (1993) and Francq and Zakoian (2000), consistency of the IOLS estimator also holds.

We shall prove the consistency of the IOLS estimator. We first show that the $N(\phi)$, $\phi \in \mathbb{B}$, is an ACM and thus has a fixed point (see Lemma 1 in Appendix). Lemma 1 guarantees that the population mapping is an ACM if $\left|\frac{\beta_1\beta_2}{1+\beta_1\beta_2}\right| < 1$.

From Lemma 1, we have that the population mapping is an ACM if the eigenvalues associated with the gradient of the population mapping evaluated at β, denoted by $V(\beta)$, have absolute values smaller than one.

$$V(\beta) = \begin{pmatrix} \frac{\beta_2}{\beta_1+\beta_2} & \frac{\beta_2\left(1-\beta_1^2\right)}{(\beta_1+\beta_2)(1+\beta_1\beta_2)} \\ \\ \frac{-\beta_2}{\beta_1+\beta_2} & \frac{-\beta_2\left(1-\beta_1^2\right)}{(\beta_1+\beta_2)(1+\beta_1\beta_2)} \end{pmatrix} \tag{2.24}$$

From Lemma 1, the two eigenvalues (λ_1, λ_2) associated with (2.24) are

given by:

$$\lambda_1 = 0 \tag{2.25}$$

$$\lambda_2 = \frac{\beta_1\beta_2}{(1+\beta_1\beta_2)} \tag{2.26}$$

If both β_1 and β_2 have the same sign, then $\beta_1\beta_2 > 0$, which leads to $|\lambda_2| < 1$. Hence, under the condition that both parameters share the same sign, we have that Lemma 1 holds. If β_1 and β_2 have different signs, then there are different combinations of parameters that violate the ACM condition, i.e., $|\lambda_2| > 1$.

To check all combinations such that Lemma 1 holds, we perform a numerical analysis using (2.26). We execute a numerical grid search on λ_2 through all possible combinations of β_1 and β_2 that satisfy Assumption 1. Figure 2.1 displays the maximum eigenvalue associated with the theoretical gradient. As a sufficient rule for Lemma 1 to hold, we have that if $|\beta_1 - \beta_2| < 1.41$ then $|\lambda_2| < 1$. In Figure 2.2, we zoom in on the previous analysis and only consider the different combinations of parameters that guarantee that $\left|\frac{\beta_1\beta_2}{(1+\beta_1\beta_2)}\right| < 1$. We show that for a large area of the graph, $|\lambda_2|$ is smaller than 0.5. This is a particularly important result, because it defines an upper bound for the number of iterations that the IOLS algorithm may take to converge. Thus, if $\kappa = 0.5$ and $T = 1000$, we would require no more than 15 iterations to achieve convergence with a precision of three decimal places. The validity of Lemma 1 is crucial to proving the consistency of the IOLS estimator. If $N(\phi)$ is an ACM, then $d(N(\phi) - N(\gamma)) \leq \kappa d(\phi - \gamma)$ holds, with $\gamma, \phi \in \mathbb{B}$, $\kappa \in [0, 1)$, and $d(.)$ being any distance function. Moreover, $N(\phi)$ will have a fixed point on

$(\mathbb{B}, d)$, as discussed in Dominitz and Sherman (2005).

We claim that the additional condition on the parameters for $N(\phi)$ to be an ACM is not very restrictive because the initial estimator we use is already consistent. Therefore, in the cases where the true data generation process is not an ACM, the IOLS algorithm will not converge and we thus adopt the consistent initial estimates. The major disadvantage in using the initial estimator is the larger variance associated with the parameter estimates.

To show the consistency of the IOLS estimator, we require three further conditions: the population and sample mapping converge uniformly in probability (Lemma 3); uniform convergence on the gradients of the mappings (Lemma 4); and the sample mapping is also an ACM (Lemma 5). Lemmas 3 and 4 are crucial to show that the sample mapping is also an ACM and thus has a fixed point denoted by $\widehat{\beta}$, such that $\widehat{N}_T\left(\widehat{\beta}\right) = \widehat{\beta}$. Proofs of the Lemmas related to the conditions above are stated in the Appendix.

To establish an asymptotic distribution of the IOLS estimator, we evoke similar conditions as those stated in Theorem 4 in Dominitz and Sherman (2005). Lemma 6 provides $\sqrt{T}$ convergence of $\widehat{\beta}^{j(T)}$ to the fixed point of the sample mapping. We show that Lemma 6 holds, provided that $j(T)$ increases at a sufficient rate of T. From the ACM definition, κ is bounded such that $\kappa \in [0, 1)$. We thus have that $\kappa^{j(T)}$ dominates $\sqrt{T}$ yielding Lemma 6 if $\frac{\ln(T)}{j} = o(1)$ holds, where j denotes the number of iterations. Define $A = [I_m - V(\beta)]^{-1}$ and $H = \text{plim}\left[\frac{X'_{-1}X_{-1}}{T}\right]$, then Theorem 1 delivers the consistency and asymptotic distribution of the IOLS estimator.

Theorem 1 *Suppose Assumptions 1 and 2 hold and* $\left|\frac{\beta_1\beta_2}{1+\beta_1\beta_2}\right| < 1$. *Then,*

i. $\left|\widehat{\beta} - \beta\right| = o_p(1)$ *as* $j(T) \longrightarrow \infty$ *with* $T \longrightarrow \infty$.

ii. $\sqrt{T}\left[\widehat{\beta} - \beta\right] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2 A H^{-1} A')$ *as* $T \longrightarrow \infty$ *and* $j(T) \longrightarrow \infty$

Note that X_{-1} is the matrix of regressors computed using the true vector of parameters β. A closed-form expression for the gradient function is given in (2.24). Estimates of β can be used to compute the empirical counterparts of both $V(\beta)$ and H.

This result holds for any starting value, provided that $\widehat{\beta}^0 \in \mathbb{B}$, which yields an interesting theoretical property of this class of estimator. In the particular case of the ARMA(1,1) model, it is enough to choose initial estimates that fulfill Assumption 1, which turns out to be very simple. It is also relevant that for item (i) in Theorem 1 to hold, no particular rate is required for $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$.

Monte Carlo simulations designed to check the theoretical asymptotic distribution for the ARMA(1,1) model deliver consistent estimates for the empirical asymptotic variances. These findings strengthen item (ii) in Theorem 1. These results are available upon request.

The extension of Theorem 1 to the VARMA class of models is theoretically straightforward, but mathematically cumbersome. The crucial point is showing that the VARMA mapping is an ACM. To this purpose, one would have to follow the steps in Lemma 1 up to (2.37). From this point onwards, the simplification of the infinite expansions into simpler functions becomes problematic. We perform numerical experiments testing whether the gradient from the VARMA mapping has eigenvalues less than one in absolute value. The results are very similar to those from the univariate case (i.e., provided that some conditions on the eigenvalues of both autoregres-

sive and moving average parameter matrices hold, we have that the OLS mapping is an ACM). This finding matches our Monte Carlo results, in which the IOLS estimator achieves convergence with different eigenvalues and system dimensions.

2.4 Monte Carlo Study

This section provides results that shed light on issues such as consistency and forecast accuracy of VARMA models estimated using the IOLS methodology. We consider three different types of Monte Carlo exercises in this section. The first exercise addresses the comparison between IOLS and MLE estimators in small- and medium-sized systems. The second exercise focuses on analyzing the consistency of the IOLS estimator. We design simulations covering different model dimensions, sample sizes, and dependencies among the variables. The last set of exercises compares the forecast performances of VARMA models estimated using the IOLS methodology with those of VAR and AR(1) models.

We define our DGP by assuming a VARMA model such as the one stated in (2.1). Following the discussion in Section 2.2, we assume that our baseline model is stable, invertible, and unique. We generate the disturbances using pseudo random standard normal in GAUSS. Uniqueness is satisfied by setting all Kronecker indices to one. This implies that our baseline model is a VARMA(1,1) as stated in (2.27), where $A_0 = M_0 = I_K$ and both A_1 and M_1 are full matrices. By setting all Kronecker indices to one, the matrix A_0 is restricted to be an identity matrix, implying that the dynamics of the VARMA model is determined by the eigenvalues of A_1

and M_1.

$$Y_t = A_1 Y_{t-1} + U_t + M_1 U_{t-1} \tag{2.27}$$

We discard the initial 500 observations to reduce dependence on initial conditions. The number of replications varies across the different Monte Carlo experiments, because of the time of computation associated to the different specifications.

We aim to design models that present highly correlated series as a way to replicate the task of forecasting macroeconomic variables. We generate the matrices A_1 and M_1 in such a way that we can control the dynamics of the process $\{Y_t\}_{t=1}^{T}$. By defining the eigenvalues of A_1 and M_1, we are able to control the persistence of the autoregressive and moving-average components, and consequently the correlation among the K variables in the system. We implement the method suggested by Camba-Mendez and Kapetanios (2004). Because A_1 and M_1 are generated in the same way, we only describe the structure referring to matrix A_1. We define $A_1 = \tilde{E}D\tilde{E}'$, where both D and $\tilde{E}$ have dimensions $(K \times K)$. We construct matrix $\tilde{E}$ in two steps. First, we generate a $(K \times K)$ matrix using pseudo random standard numbers, which is then orthonormalized by applying the Gram-Schmidt methodology. The key factor for specifying the dynamic of matrices A_1 and M_1, and consequently for the process $\{Y_t\}_{t=1}^{T}$, is the matrix D. We specify D as being a quasi upper triangular matrix, where all the (2×2) blocks on the diagonal have eigenvalues equal to those specified for the simulation. When K is an odd number, the remaining last element of the diagonal of matrix Λ is made equal to the last eigenvalue of the system. Originally, we set the remaining non-zero values of D to one. As the number

of variables in our baseline VARMA model increases, the volatility of the process $\{Y_t\}_{t=1}^{T}$ increases very quickly, which does not correspond to real economic data. To reduce the volatility without changing the assigned eigenvalues of matrix A_1, we shrink all the off-diagonal, non-zero elements of matrix D towards zero. We implement this shrinkage strategy for both matrix A_1 and matrix M_1. When the system size is large, the diagonal elements of both the A_1 and M_1 matrices are close to the values assigned for the eigenvalues, whereas the off-diagonal elements approach zero.

The first set of Monte Carlo simulations addresses the relative performance of the IOLS estimator compared to the MLE estimator. To assess this comparison, we report the relative mean squared error (RelMSE), computed as RelMSE $= \frac{MSE_{IOLS}}{MSE_{MLE}}$. A RelMSE less than one indicates that the IOLS estimator performs better than MLE, in terms of MSE. To assess computational demand and rate of convergence, we report the relative time of computation (RT) and the percentage of failure (%F) computed within all replications. In the first exercise, we estimate a system with three variables ($K = 3$). We assign eigenvalues equal to 0.5 for both autoregressive and moving average parameter matrices; and we consider samples of 50, 100, 150, 200 and 400 observations.

Table 2.1 displays results showing that the IOLS performs better (in terms of the MSE) than the MLE[2] estimator when the sample size is either 50 or 100. We find that IOLS provides more accurate estimates for both autoregressive and moving average parameters, showing average gains of 17% and 41% when $T = 50$ and $T = 100$, respectively. It is important

[2]We use the CML - Maximum Likelihood Estimation with General Nonlinear Constraints on Parameters - package in GAUSS. We also perform simulations where the initial values of the MLE algorithm are defined as being the IOLS estimates, however, the overall picture remains the same.

to point out that in this small dataset experiment, we are estimating 18 parameters, which turns out to be quite demanding for both estimators when $T = 50$. With regard to the time of computation, we also show that the MLE estimator takes much longer to be computed (average of 677 times longer than the IOLS estimator when $T = 100$). We design an alternative Monte Carlo exercise where we compare the IOLS estimator with a constrained version of the MLE estimator. We impose the constraint on the eigenvalues of matrix M_1 because we identify them as the major cause for non-convergence. In spite of achieving convergence in 100% of replications, the constrained MLE algorithm is not a feasible alternative for medium and large datasets due to its computational demand. We do not report the results of these simulations, because they follow the same pattern (in terms of the RelMSE) as do the unconstrained Monte Carlo simulations.

To assess the feasibility of the MLE estimator when applied to larger systems, we design Monte Carlo experiments considering VARMA models with eight and ten variables at different sample sizes: $T = 100$, $T = 150$, $T = 200$ and $T = 400$. The number of replications were truncated to small numbers following the computation demand[3] on obtaining the MLE estimator. Therefore, the results depicted in Table 2.2 do not carry any statistical properties, but they do shed light on the limitations of the MLE estimator when dealing with high-dimensional systems. From Table 2.2, we show that for both $K = 8$ and $K = 10$ the MLE fails more than the IOLS estimator for $T \leq 200$. When $T = 400$ and $K = 8$, the MLE estimator presents a lower rate of failure than the IOLS estimator, however its

[3]Simulations were carried out in two dedicated UNIX servers with seven and eleven processors respectively.

computation time is much higher (RT = 469.8). When $K = 10$, the picture becomes even less favorable to the MLE estimator. Table 2.2 displays results showing that this estimator becomes unfeasible for this system dimension, supporting the conclusion that MLE is not feasible for medium and large datasets. These findings motivate us to implement the IOLS estimator for forecasting macroeconomic variables using medium and large datasets.

The second set of Monte Carlo exercises addresses the consistency of the IOLS estimator when considering medium and large datasets. We report the root mean squared error (RMSE) of the parameter estimates as an assessment of consistency. To disentangle the difference in performance from the replications that converged and the ones that did not, we report results in two ways: first, within all replications; second, considering only the ones that converged (denoted with the superscript c).

We discuss four different model specifications in this section. We design the first three models with $K = 10$, and the fourth specification with $K = 20$. When $K = 10$, eigenvalues of both parameters matrices are set to 0.3, 0.8 and mixed, in the first, second and third models respectively. When $T = 1000$ we perform 2000 replications. For all the remaining sample sizes we set the number of replications equal to 10000. Tables 2.3, 2.4 and 2.5[4] report results for samples of 150, 200, 400 and 1000 observations[5]. As we expect, the parameters from the autoregressive matrix converge very fast to the true value, whereas the moving average parameters require more observations (larger T) to converge to the true value. Comparing the results from the

[4]Due to space limitations, we do not report the estimates of all elements of matrices A_1 and M_1, but only their diagonal elements.

[5]We do not construct simulations with $T < 150$, because of the lack of the degrees of freedom following the estimation of the VAR(p) in the first step of the IOLS algorithm.

two specifications (especially the case where $T = 200$), we conclude that the eigenvalues of M_1 play a very important role in determining the rate at which the moving average parameters converge to the true value. Hence, the higher the eigenvalues of M_1 are, the slower the rate of convergence of the moving average parameters is. Moreover, the eigenvalues associated M_1 determine the percentage of failure of the IOLS estimator and therefore the quality of the estimates. This will play an important role on the forecast accuracy of the models estimated with the IOLS estimator. Finally, results from both models corroborate our claim that IOLS estimates are consistent, since their RMSE's decay towards zero as sample size increases. Finally, apart from the case where the eigenvalues associated to the parameter matrices are equal to 0.8, the IOLS rate of failure is around 30% for $T = 400$. This turns out to be quite robust, following the fact that we estimate 200 parameters in this specification.

We design the third exercise to have $K = 20$ with eigenvalues from both autoregressive and moving average parameter matrices equal to 0.6. It is important to stress that since we set all Kronecker indices equal to one, we estimate 800 parameters in this simulation. Table 2.6 reports results from different samples: $T = 200$, $T = 400$, and $T = 1000$ observations. We find a similar pattern for the evolution of RMSE compared with that of lower dimensional models: the autoregressive parameters converge to the true values very fast, whereas the IOLS estimates from the moving average parameters require a larger sample size. Following the results presented in Table 2.6, we conclude that the IOLS estimation procedure is feasible and consistent, even for large systems. Moreover, we expect that as the number of free parameters decreases (following the Echelon form transformation),

the rate of failure also will decline.

We focus the last set of Monte Carlo exercises on assessing the forecast accuracy of VARMA models estimated using the IOLS methodology. Our main objectives are: understanding the tradeoff, in terms of forecast accuracy, between sample size and system dimension; and assessing the difference in forecast performance from the replications that converged with respect to the ones that did not. We report results considering four different model specifications. The first three models have $K = 10$, but they differ with respect to the eigenvalues we assign for A_1 and M_1 (0.3, 0.8, and mixed eigenvalues, respectively). The fourth model has $K = 20$ and all eigenvalues equal to 0.6. We report results in terms of the relative mean squared forecast error (RelMSFE). RelMSFEs are computed as the ratio of the mean squared forecast error (MSFE) of the VARMA model and the MSFE of the competitor model. Therefore, we have that a RelMSFE less than one implies that the VARMA model outperforms the competitor model. We compare VARMA models against VAR(p) (where p is chosen according to the AIC criterion) and AR(1). We report the relative measures computed using only the replications that converged (denoted with the superscript c) and using all replications. Due to the large number of variables, we report the mean of the RelMSFE computed within all variables in the system. We report results considering forecast horizons up to twelve-steps-ahead.

Tables 2.7, 2.8, and 2.9 show that when is T small ($T = 150$, $T = 200$) and the IOLS algorithm does not converge, VARMA is easily outperformed by the AR(1) and VAR models. Still considering small samples, we also observe that when the eigenvalues associated to A_1 and M_1 are either low

or mixed, the difference in performance of the VARMA model with respect to the AR(1) model when the IOLS estimator does not converge is much smaller than in the case with high eigenvalues. This better forecast performance has two determinants: first, models with lower eigenvalues associated to the parameter matrices have a higher rate of convergence in small samples than models with high eigenvalues, yielding a higher share of converged iterations that contributes to a lower RelMSFE. Second, the initial estimator appears to do a worse job in forecasting models with high eigenvalues. This contributes to the very poor performance of the VARMA model when $T = 150$ as observed in Table 2.8. Combining this result with those reported in Tables 2.3 and 2.4, we conclude that systems with lower eigenvalues associated with the parameter matrices achieve consistency at a faster rate, and thus deliver more accurate forecasts than systems with eigenvalues close to one. By shifting our analysis for the cases where $T = 400$, we report results showing that VARMA models estimated with IOLS outperform both VAR and AR(1) models. In particular, we note that the RelMSFE measures computed using only the replications that converged are quite stable through all the sample sizes. This implies that when the IOLS converges, the forecast accuracy obtained from VARMA models outperforms both the AR(1) and VAR models in all sample sizes. Furthermore, our results show that differences in performance of VARMA model against the competitors models are quite smooth across all the forecast horizons. In particular, the first-step-ahead forecast tends to be the one that varies the most, following changes in sample size. Table 2.10 reports results considering the case where $K = 20$. These results present the same pattern as the ones displayed in Tables 2.7, 2.8, and 2.9. The most

significant difference lies on the number of observations the IOLS estimator needs to improve the convergence rate. As in this setup we estimate 800 parameters, it is necessary 400 observations for VARMA models outperform the AR(1) model considering both measures. This result will be crucial in defining the sample size we will use in the empirical application in Section 2.5.

2.5 Empirical Application

In this section, we analyze the competitiveness of VARMA models estimated with the IOLS procedure to forecast macroeconomic variables. Our aim is to forecast three key macroeconomic variables: industrial production (IPS10), interest rate (FYFF), and CPI inflation (PUNEW). We asses VARMA forecast performance under different system dimensions and forecast horizons.

2.5.1 Data and Setup

We use US monthly data from the Stock and Watson (2005) dataset, which runs from 1959:1 through 2003:12. We do not use all the available series from this dataset; as in Carriero, Kapetanios, and Marcellino (2011), we use 52 macroeconomic variables that represent the main categories of economic indicators. From the 52 selected variables, we work with three system dimensions: $K = 10$, $K = 20$, and $K = 40$. We construct five different datasets (one to five) for each system size. We restrict the maximum number of variables in the system to $K = 40$ because of computational constraints. There is no particular rule to select the variables within the

entire group of 52; however, we try to keep a balance among the three main categories of data: real economy, money and prices, and financial market. The series are transformed, as in Carriero, Kapetanios, and Marcellino (2011), in such a way that they are approximately stationary. We report the dataset details in Appendix 1, Table 2.11.

So far, we have assumed that the Kronecker Indices are all known, which implies that any general VARMA model can be written in Echelon form by applying the procedure described by equations (2.5), (2.6), (2.7), (2.8) and (2.9). When one is dealing with empirical data, however, the true DGP is unknown as, consequently, are the Kronecker indices. The task of defining the Kronecker Indices is a crucial step in our forecast analysis. We specify the Kronecker indices according to three different algorithms.

The first algorithm that we adopt is the Hannan-Kavalieris (HK) procedure, as discussed in Hannan and Kavalieris (1984b), Hannan and Kavalieris (1984a) and Lütkepohl (2007). We stress the importance of this algorithm because we construct the next two alternative methodologies based on the HK procedure. The HK algorithm consists of minimizing information criterion denoted by $C(\mathbf{p})$, given different alternative specifications of $\mathbf{p}$, where $\mathbf{p}$ is a $(K \times 1)$ vector of Kronecker indices. We can split the procedure into two steps. First, we start the procedure by exogenously defining the maximum value that the Kronecker indices may assume, which is denoted by $p_{\max}$. Following that, we estimate different VARMA models, assigning all elements of $\mathbf{p}$ to be equal to $p_{\max}, p_{\max} - 1, ..., 1$, successively. We choose $\mathbf{p}$ that minimizes the criterion $C(\mathbf{p})$, denoting the vector of Kronecker indices as $p^{(1)}$, where $1 \leq p^{(1)} \leq p_{\max}$. In the second step of the HK algorithm, we define the Kronecker indices that will be

used to estimate the model. This strategy requires K evaluations, since we aim to define $p_k = \hat{p}_k$, for all $k = 1, ..., K$ by minimizing the information criterion successively. The first evaluation requires the estimation of the model varying the K^{th} Kronecker index from $p_K = 0$ to $p_K = p^{(1)}$. We choose the $\hat{p}_K$ associated with the lowest value of $C(\mathbf{p})$. At the end of this first evaluation, we have $\mathbf{p} = \left(p_1 = p^{(1)}, p_2 = p^{(1)}, ..., p_K = \hat{p}_K\right)'$. We repeat the procedure for all of the remaining Kronecker indices. Therefore, the k^{th} evaluation results in the following vector of Kronecker indices: $\mathbf{p} = \left(p_1 = p^{(1)}, ..., p_{k-1} = p^{(1)}, p_k = \hat{p}_k, \hat{p}_{k+1}, ..., \hat{p}_K\right)'$. The information criterion is chosen to be the Schwarz criterion (SC), as it delivers consistent estimates of the Kronecker indices when U_t is a strong white noise process and the VARMA model is invertible and stable.

In the spirit of Kapetanios, Labhard, and Price (2006), we design the second algorithm (denoted as MT) to recover a specification that produces the most accurate forecast (lowest MSFE). The algorithm consists of finding the Kronecker index that minimizes the trace of the out-of-sample MSFE of the key macroeconomic variables of interest. We set up the algorithm in a very similar way to the HK procedure: we repeat the first and second steps of the HK procedure, but swap the SC criterion for the trace of the (3×3) upper block of the MSFE matrix as the criterion we need to minimize. Comparing our second algorithm with the HK procedure, we expect that the former will deliver a more accurate forecast of the key macroeconomic variables, because SC criterion is a function of the entire covariance matrix of the residuals.

Our last algorithm (denoted OZ) consists of setting all Kronecker indices equal to zero, except for the ones related to the key macroeconomic

variables that we want to forecast. By ordering the three key variables on the top of the system and implementing this specification, we force the matrix $A_0^{-1}A_1$ to have non-zero elements only in the first three columns, whereas the matrix $A_0^{-1}M_1$ remains full. We are thus able to capture the AR(1) feature usually present in macroeconomic variables, as well as allow for a rich dynamic in the lagged shocks. In spite of its simplicity, we show that this algorithm is quite competitive and robust compared with the other specifications.

2.5.2 Results

We report forecast results considering the three algorithms discussed in the previous subsection. We consider as competitor models the following specifications: VAR(1), AR(1), and AR(1) with constant (denoted as AR(1)‡). As in the Monte Carlo simulations, we compare different models using the out-of-sample RelMSFE, where the numerator always contains the MSE computed from the VARMA model. Therefore, a RelMSFE less than one indicates that the VARMA model outperforms the competitor model. We compare the prediction accuracy among the different models using the Diebold and Mariano (1995) test. We set the sample size equal to 470 observations in all exercises, because we conclude in Section 2.4 that VARMA models require a larger sample size to outperform AR(1) specifications. We perform 50 out-of-sample forecasts considering two different horizons: first- (Hor:1) and fourth- (Hor:4) step-ahead. The Kronecker indices in all algorithms were set to present a maximum value of one, which implies that all VARMA models are VARMA(1,1). We subtract the sample mean at an initial stage, which implies that all the models use mean-

adjusted variables. By doing so, we do not require any constants in our baseline model.

Table 2.12 reports the results of systems with 10 variables ($K = 10$) estimated with the HK, MT, and OZ algorithms. For each of the algorithms, we report results from five different datasets. We do not report results from the OZ algorithm using dataset four, due to a lack of stability in the initial estimator. Considering the HK and OZ algorithms, we observe that the VARMA long-horizon forecasts tend to outperform those of competitors; whereas VARMA first-step-ahead forecasts are usually beaten, especially those for the FYFF variable. In particular, we point out decent performances of the HK algorithm in datasets three and four. The MT algorithm, as expected, delivers more accurate forecasts on both horizons for all the variables. For the first-step-ahead forecast, for instance, we report gains up to 35%, 33%, and 13% with respect to the AR(1) model for IPS10, FYFF, and PUNEW, respectively.

Table 2.13 displays results for large datasets ($K = 20$). We show that the MT algorithm produces very accurate fourth-step-ahead forecasts (mostly for dataset two) with gains up to 31% in IPS10. Considering the first-step-ahead forecast, VARMA does not outperform either the AR(1) model or the VAR(1) model. As discussed in Carriero, Kapetanios, and Marcellino (2011), as K gets large, AR(1) specification becomes extremely hard to beat for short horizons. It is also important to point out the good performance of the VAR(1) models, which may be justified by the large sample size ($T = 470$). Regarding the HK algorithm, VARMA performs well with datasets three and four for the fourth-step-ahead forecast, but these results are not statistically significant.

Table 2.14 reports results for the models with large datasets ($K = 40$). In accordance with previous findings, we observed that VARMA models do a better job at forecasting longer horizons in all three algorithms. From Table 2.14, we show that the HK algorithm delivers good results when forecasting the fourth-step ahead. We report gains from 12% to 5% in dataset four. Surprisingly, HK specification outperforms the AR(1) benchmark on the first-step-ahead forecast with dataset four, with gains of 7% for IPS1. We conclude that, by considering all K variables in the system to determine the Kronecker indices, the HK algorithm does not systematically outperform the AR(1) model, indicating that more restrictions may be needed. Considering the MT algorithm in Table 2.14, we find that VARMA models outperform the VAR specification for the PUNEW variable across all the different datasets. We report gains up to 43% in both datasets 1 and 5. Comparing the VARMA performance with the AR(1) results, we find that all RelMSFE are very close to one, indicating that both models perform equally well. In special, we find an outstanding gain of 22% for the FYFF variable in dataset four. Finally, the OZ algorithm outperforms the AR(1) benchmark for at least one of the three key macroeconomic variables, when considering longer horizons, in all datasets but dataset four. Although results are not significant, we report gains up to 11% for both IPS10 and FYFF.

Comparing the results obtained with the OZ, HK, and MT algorithms, we conclude that the OZ algorithm performs worse than the HK and MT algorithms in all system dimensions but when $K = 40$. In addition to that, we find that forecast results are very sensitive to Kronecker indexes specification, implying that specification plays a decisive role on improv-

ing forecast accuracy in VARMA models. Therefore, we conclude that the VARMA class of models is able to incorporate the information presented in medium and large datasets while also attenuating the "curse of dimensionality ".

To sum up the results of this section, we show that VARMA models outperform both AR(1) and VAR(1) models for the fourth-step-ahead forecast. This finding is especially present for the MT algorithm and when $K = 10$ and $K = 20$. Considering the first-step-ahead forecast, VARMA models outperform the VAR(1) specification in different algorithms and system sizes. When it turns to compete against the AR(1) model, however, VARMA models are not able to beat this competitor apart from some specifications when $K = 10$. When considering $K = 40$, we find that the OZ algorithm delivers more robust results in the fourth-step-ahead forecast, outperforming the AR(1) model in at least one variable in all but one dataset. This result supports the previous findings in this literature (see Carriero, Kapetanios, and Marcellino (2011), De Mol, Giannone, and Reichlin (2006), Banbura, Giannone, and Reichlin (2007), Carriero, Kapetanios, and Marcellino (2008), among others), that, by imposing restrictions on the parameter matrices, we are able to improve forecast accuracy when dealing with large datasets. For $K = 20$ and $K = 10$, however, we find mixed evidences, leading to the conclusion that the VAR(1) class of models remains quite powerful under these system dimensions, especially if the sample size is reasonably large. In addition, the forecast performance depends heavily on the algorithm implemented to define the Kronecker indices. Although those procedures are very time consuming, they are important and must be undertaken for all datasets. To conclude, we highlight

the good performance of VARMA models in medium datasets ($K = 10$), where the IOLS algorithm is able to outperform both VAR and AR(1) models in different datasets.

2.6 Conclusion

This chapter addresses the issue of forecasting key macroeconomic variables using medium and large datasets. We propose the use of VARMA models as a feasible framework for this task. We overcome the natural difficulties in estimating medium- and high-dimensional VARMA models with the MLE framework by adopting the IOLS estimator.

We establish the consistency and asymptotic distribution for the IOLS estimator by considering the univariate ARMA(1,1) model and we argue that these results can be extended to the multivariate case. Our Monte Carlo exercises corroborate our theoretical findings, and support their validity to the VARMA case. It is also important to point out that our theoretical results are obtained under very weak assumptions. This qualifies the IOLS estimator to cover specifications similar to the ones covered by the quasi-maximum likelihood estimator. Our Monte Carlo study shows that the IOLS estimator is feasible and consistent in high-dimensional systems. Furthermore, we also report results showing that the IOLS estimator outperforms the MLE, in terms of mean squared error, when T is small. The empirical results show that VARMA models perform better than VAR(1) and AR(1) models for different system sizes. We find that VARMA models do a better job at forecasting longer horizons. In particular, the models we specify using the MT algorithm produce the most accurate results. We also conclude that, as system dimensionality increases, the specification of Kronecker indices tends to play a more important role in improving forecast accuracy.

Finally, based on both the Monte Carlo exercises and the empirical application, we conclude that VARMA models, which are estimated using

the IOLS methodology, produce competitive forecasts and qualify as valid alternatives for forecasting key macroeconomic variables such as industrial production, inflation, and interest rates.

2.7 Appendix

Lemma 1 *Suppose Assumptions 1, 2 hold and* $\left|\frac{\beta_1\beta_2}{1+\beta_1\beta_2}\right| < 1$. *Then, there exists an open ball centered at* β *with closure* $\mathbb{B}$, *such that the mapping* $N(\phi)$ *is an ACM on* $(\mathbb{B}, d)$, *with* $\phi \in \mathbb{B}$.

Proof of Lemma 1: We mirror our proof in Lemma 5 in Dominitz and Sherman (2005). By Taylor expansion, we rewrite $N(\phi)$ around γ, with $\phi, \gamma \in \mathbb{B}$. There also exists a ϕ^* located in the segment line between ϕ and γ, such that $|N(\phi) - N(\gamma)| = |V(\phi^*)[\phi - \gamma]|$ holds. Combining the two results, we are in a position to define a bound that is function of the gradient of the population mapping evaluated on β.

$$\begin{aligned} |N(\phi) - N(\gamma)| = |V(\phi^*)[\phi - \gamma]| \leq |V(\beta)[\phi - \gamma]| + \\ + |[V(\phi^*) - V(\beta)][\phi - \gamma]| + o_p(|\phi - \gamma|) \end{aligned} \tag{2.28}$$

From their result, it suffices to show that the maximum eigenvalue of $V(\beta)$ is less than one in absolute value. To this purpose, we define $V\left(\beta^{j(T)}\right) = \bigtriangledown_{\beta^{j(T)}} N\left(\beta^{j(T)}\right)$ as the gradient from the population mapping on the $(j+1)^{th}$ iteration.

$$V\left(\beta^{j(T)}\right) = \begin{bmatrix} \frac{\partial \beta_1^{j(T)+1}}{\partial \beta_1^{j(T)}} & \frac{\partial \beta_1^{j(T)+1}}{\partial \beta_2^{j(T)}} \\ \\ \frac{\partial \beta_2^{j(T)+1}}{\partial \beta_1^{j(T)}} & \frac{\partial \beta_2^{j(T)+1}}{\partial \beta_2^{j(T)}} \end{bmatrix} \tag{2.29}$$

Using the partitioned regression result, we obtain individual expressions for the OLS estimates of β obtained from the population mapping at each

iteration.

$$\beta_1^{j(T)+1} = \mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1} \mathbb{E}\left[\frac{1}{T}Y'_{-1}\left[Y - U^j_{-1}\beta_2^{j(T)+1}\right]\right] \tag{2.30}$$

$$\begin{aligned}\beta_2^{j(T)+1} = &\left[\mathbb{E}\left[\frac{U^{j\prime}_{-1}U^j_{-1}}{T}\right] - \mathbb{E}\left[\frac{U^{j\prime}_{-1}Y_{-1}}{T}\right]\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1}\mathbb{E}\left[\frac{Y'_{-1}U^j_{-1}}{T}\right]\right]^{-1} \times \\ &\left[\mathbb{E}\left[\frac{U^{j\prime}_{-1}Y}{T}\right] - \mathbb{E}\left[\frac{U^{j\prime}_{-1}Y_{-1}}{T}\right]\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1}\mathbb{E}\left[\frac{Y'_{-1}Y}{T}\right]\right]\end{aligned} \tag{2.31}$$

Using the invertibility condition to express estimates of the lagged disturbances as in (2.22), we have:

$$\begin{aligned}\frac{\partial\beta_1^{j(T)+1}}{\partial\beta_1^{j(T)}} = &\left[\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1} \times \right. \\ &\left.\mathbb{E}\left[\frac{1}{T}Y'_{-1}\left[\left(1+\beta_2^{j(T)}L\right)^{-1}Y_{-2}\right]\right]\right]\beta_2^{j(T)+1} - \\ &\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1}\mathbb{E}\left[\frac{Y'_{-1}U^j_{-1}}{T}\right]\frac{\partial\beta_2^{j(T)+1}}{\partial\beta_1^{j(T)}}\end{aligned} \tag{2.32}$$

Evaluating (2.32) on the true vector of parameters β, the first element of (2.29) reduces to:

$$\left.\frac{\partial\beta_1^{j(T)+1}}{\partial\beta_1^{j(T)}}\right|_\beta = \left(\frac{1}{\sigma_y^2}\right)\left[\sum_{i=0}^{\infty}(-\beta_2)^i\gamma_{1+i}\right]\beta_2 - \left(\frac{\sigma_u^2}{\sigma_y^2}\right)\left[\left.\frac{\partial\beta_2^{j(T)+1}}{\partial\beta_1^{j(T)}}\right|_\beta\right] \tag{2.33}$$

where $\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right] = \sigma_y^2 = \frac{\left(1+\beta_2^2+2\beta_1\beta_2\right)\sigma_u^2}{\left(1-\beta_1^2\right)}$ is the variance of the ARMA(1,1) process, $\mathbb{E}\left[\frac{Y'_{-1}U_{-1}}{T}\right] = \sigma_u^2$ is the variance of the disturbances and $\mathbb{E}\left[\frac{Y'Y_{-l}}{T}\right] = \gamma_l = \beta_1^{l-1}\left(\beta_1\sigma_y^2 + \beta_2\sigma_u^2\right)$ is the autocovariance of lag l.

Similarly to (2.32), the second element in the first row of (2.29) is:

$$\begin{aligned}\frac{\partial \beta_1^{j(T)+1}}{\partial \beta_2^{j(T)}} = &\left[\mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1}\mathbb{E}\left[\frac{1}{T}Y'_{-1}\left[\left(1+\beta_2^{j(T)}L\right)^{-1}U^j_{-2}\right]\right]\right]\times \\ &\beta_2^{j(T)+1} - \mathbb{E}\left[\frac{Y'_{-1}Y_{-1}}{T}\right]^{-1}\mathbb{E}\left[\frac{Y'_{-1}U^j_{-1}}{T}\right]\frac{\partial \beta_2^{j(T)+1}}{\partial \beta_2^{j(T)}}\end{aligned} \tag{2.34}$$

Evaluating (2.34) on the true vector of parameters β, the second element in the first row of (2.29) reduces to (2.35), with $\mathbb{E}\left[\frac{Y'U_{-l}}{T}\right] = \gamma_l^* = \beta_1^{l-1}\left[\sigma_u^2\left(\beta_1+\beta_2\right)\right]$.

$$\left.\frac{\partial \beta_1^{j(T)+1}}{\partial \beta_2^{j(T)}}\right|_\beta = \left(\frac{1}{\sigma_y^2}\right)\left[\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma^*_{1+i}\right]\beta_2 - \left(\frac{\sigma_u^2}{\sigma_y^2}\right)\left[\left.\frac{\partial \beta_2^{j(T)+1}}{\partial \beta_2^{j(T)}}\right|_\beta\right] \tag{2.35}$$

Computing the elements in the second row of (2.29) in a similar manner as in (2.33) and (2.35) we have:

$$\begin{aligned}\left.\frac{\partial \beta_2^{j(T)+1}}{\partial \beta_1^{j(T)}}\right|_\beta = &-2\left[\sigma_u^2 - \frac{\left(\sigma_u^2\right)^2}{\sigma_y^2}\right]^{-2}\times \\ &\left[\gamma_1^* - \frac{\sigma_u^2\gamma_{-1}}{\sigma_y^2}\right]\left[\left(\frac{\sigma_u^2}{\sigma_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i}\right)\right] + \\ &\left[\sigma_u^2 - \frac{\left(\sigma_u^2\right)^2}{\sigma_y^2}\right]^{-1}\left[-\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{2+i}\right) + \right. \\ &\left.\left(\frac{\gamma_1}{\sigma_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i}\right)\right]\end{aligned} \tag{2.36}$$

$$\left.\frac{\partial \beta_2^{j(T)+1}}{\partial \beta_2^{j(T)}}\right|_{\beta} = -2\left[\sigma_u^2 - \frac{\left(\sigma_u^2\right)^2}{\sigma_y^2}\right]^{-2}\left[\gamma_1^* - \frac{\sigma_u^2\gamma_{-1}}{\sigma_y^2}\right]\times$$
$$\left[\left(\frac{\sigma_u^2}{\sigma_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i}^*\right)\right] + \left[\sigma_u^2 - \frac{\left(\sigma_u^2\right)^2}{\sigma_y^2}\right]^{-1}\times \tag{2.37}$$
$$\left[-\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{2+i}^*\right) + \left(\frac{\gamma_1}{\sigma_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i}^*\right)\right]$$

From (2.33), (2.35), (2.36) and (2.37) and using the fact that $\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i} = \frac{\beta_1\sigma_y^2+\beta_2\sigma_u^2}{1+\beta_1\beta_2}$, $\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{2+i} = \frac{\beta_1\left(\beta_1\sigma_y^2+\beta_2\sigma_u^2\right)}{1+\beta_1\beta_2}$, $\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{1+i}^* = \frac{(\beta_1+\beta_2)\sigma_u^2}{1+\beta_1\beta_2}$ and $\sum_{i=0}^{\infty}\left(-\beta_2\right)^i\gamma_{2+i}^* = \frac{\left(\beta_1+\beta_2\right)\beta_1\sigma_u^2}{1+\beta_1\beta_2}$, we have that (2.29) evaluated at β, denoted for simplicity as $V\left(\beta\right)$, reduces to:

$$V\left(\beta\right) = \begin{pmatrix} \frac{\beta_2}{\beta_1+\beta_2} & \frac{\beta_2\left(1-\beta_1^2\right)}{(\beta_1+\beta_2)(1+\beta_1\beta_2)} \\ \\ \frac{-\beta_2}{\beta_1+\beta_2} & \frac{-\beta_2\left(1-\beta_1^2\right)}{(\beta_1+\beta_2)(1+\beta_1\beta_2)} \end{pmatrix} \tag{2.38}$$

Note that (2.38) does not depend on σ_u^2, implying that Lemma 1 holds for any value assigned to the variance of the disturbances. The gradient of the population mapping in (2.38) has two eigenvalues: λ_1 and λ_2. These eigenvalues solve the following quadratic equation:

$$\lambda^2 + \left[-\left(\frac{\beta_2}{\beta_1+\beta_2}\right) + \left(\frac{\beta_2\left(1-\beta_1^2\right)}{\left(\beta_1+\beta_2\right)\left(1+\beta_1\beta_2\right)}\right)\right]\lambda = 0 \tag{2.39}$$

$$\lambda_1 = 0 \tag{2.40}$$

$$\lambda_2 = \frac{\beta_1\beta_2}{\left(1+\beta_1\beta_2\right)} \tag{2.41}$$

By solving (2.39), we have that (2.40) and (2.41) are the two eigenvalues associated with (2.38). Since $\lambda_1 = 0$, we only need to show that $|\lambda_2| < 1$ to prove that the population mapping is an ACM. Figure 2.1 displays $|\lambda_2|$ computed with different combinations of β_1 and β_2 such that Assumption 1 is satisfied, whereas 2.2 only shows the different combinations of parameters such that $|\lambda_2| < 1$. Combining these two numerical analyzes with the result in (2.41), we prove Lemma 1.

Lemma 2 *Suppose Assumptions 1 and 2 hold. Then, as* $T \longrightarrow \infty$, $\widehat{N}_T(\phi)$ *is stochastically equicontinuous.*

Proof of Lemma 2: We prove Lemma 2 by establishing the Lipschitz condition of $\widehat{N}_T(\phi)$ similarly as in Lemma 2.9 in Newey and McFadden (1994). We need to show that $\left\|\widehat{V}_T(\phi)\right\| = O_p(1)$ for all $\phi \in \mathbb{B}$, where $\widehat{V}_T(\phi)$ is the sample counterpart of (2.29). To this purpose, we first bound the norm of the difference of the sample mapping evaluated at different points as:

$$\left\|\widehat{N}_T(\phi) - \widehat{N}_T(\gamma)\right\| \leq \left\|\widehat{V}_T(\phi^*)\right\| \|\phi - \gamma\| \tag{2.42}$$

where $\|.\|$ accounts for the Euclidean norm, $\phi, \gamma, \phi^* \in \mathbb{B}$ and $\phi^* = (\phi_1^*, \phi_2^*)'$ lies on the segment line between ϕ and γ. The second step consists of computing the sample gradient. Note that we need to define $\widehat{V}_T(\phi^*)$ in a generic way such that it can be evaluated at any vector of estimates on any possible iteration. Using the same steps as in Lemma 1, the elements of $\widehat{V}_T\left(\beta^{j(T)}\right)$ evaluated at ϕ^* resume to:

$$\left.\frac{\partial \widehat{\beta}_1^{j(T)+1}}{\partial \widehat{\beta}_1^{j(T)}}\right|_{\phi^*} = \left(\frac{1}{\widehat{\sigma}_y^2}\right)\left[\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i \widehat{\gamma}_{1+i}\right]\phi_2^* - \left(\frac{\widehat{\zeta}_u^2}{\widehat{\sigma}_y^2}\right)\left[\left.\frac{\partial \widehat{\beta}_2^{j(T)+1}}{\partial \widehat{\beta}_1^{j(T)}}\right|_{\phi^*}\right] \tag{2.43}$$

$$\left.\frac{\partial\widehat{\beta}_1^{j(T)+1}}{\partial\widehat{\beta}_2^{j(T)}}\right|_{\phi^*} = \left(\frac{1}{\widehat{\sigma}_y^2}\right)\left[\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\delta}_{1+i}\right]\phi_2^* - \left(\frac{\widehat{\delta}_0}{\widehat{\sigma}_y^2}\right)\left[\left.\frac{\partial\beta_2^{j(T)+1}}{\partial\widehat{\beta}_2^{j(T)}}\right|_{\phi^*}\right] \tag{2.44}$$

$$\begin{aligned}\left.\frac{\partial\widehat{\beta}_2^{j(T)+1}}{\partial\widehat{\beta}_1^{j(T)}}\right|_{\phi^*} = -2&\left[\widehat{\zeta}_u^2 - \frac{\left(\widehat{\delta}_0^2\right)^2}{\widehat{\sigma}_y^2}\right]^{-2}\left[\widehat{\delta}_1 - \frac{\widehat{\delta}_0^2\widehat{\gamma}_{-1}}{\widehat{\sigma}_y^2}\right]\times \\ &\left[-\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\xi}_{1+i} + \left(\frac{\widehat{\delta}_0^2}{\widehat{\sigma}_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\gamma}_{1+i}\right)\right] + \\ &\left[\widehat{\zeta}_u^2 - \frac{\left(\widehat{\delta}_0^2\right)^2}{\widehat{\sigma}_y^2}\right]^{-1}\left[-\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\gamma}_{2+i}\right) + \right. \\ &\left.\left(\frac{\widehat{\gamma}_1}{\widehat{\sigma}_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\gamma}_{1+i}\right)\right]\end{aligned} \tag{2.45}$$

$$\begin{aligned}\left.\frac{\partial\widehat{\beta}_2^{j(T)+1}}{\partial\widehat{\beta}_2^{j(T)}}\right|_{\phi^*} = -2&\left[\widehat{\zeta}_u^2 - \frac{\left(\widehat{\delta}_0^2\right)^2}{\widehat{\sigma}_y^2}\right]^{-2}\left[\widehat{\delta}_1 - \frac{\widehat{\delta}_0^2\widehat{\gamma}_{-1}}{\widehat{\sigma}_y^2}\right]\times \\ &\left[\left(\frac{\widehat{\delta}_0^2}{\widehat{\sigma}_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\delta}_{1+i}\right)\right] + \left[\widehat{\zeta}_u^2 - \frac{\left(\widehat{\delta}_0^2\right)^2}{\widehat{\sigma}_y^2}\right]^{-1}\times \\ &\left[-\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\delta}_{2+i}\right) + \left(\frac{\widehat{\gamma}_1}{\widehat{\sigma}_y^2}\right)\left(\sum_{i=0}^{\infty}\left(-\phi_2^*\right)^i\widehat{\delta}_{1+i}\right)\right]\end{aligned} \tag{2.46}$$

where $\widehat{\zeta}_u^2 = \frac{1}{T}\sum_{t=1}^{T}u_t^j\widehat{u}_t^j$, $\widehat{\delta}_0 = \frac{1}{T}\sum_{t=1}^{T}y_t\widehat{u}_t^j$, $\widehat{\delta}_l = \frac{1}{T}\sum_{t=1}^{T}y_t\widehat{u}_{t-l}^j$, $\widehat{\xi}_l = \frac{1}{T}\sum_{t=1}^{T}y_{t-l}\widehat{u}_t^j$, $\widehat{\sigma}_y^2 = \frac{1}{T}\sum_{t=1}^{T}y_t^2$ and $\widehat{\gamma}_l = \frac{1}{T}\sum_{t=1}^{T}y_ty_{t-l}$. These quantities are all averages, and hence as $T \longrightarrow \infty$, they converge to their population counterparts. It is important to remark on two distinct results: first, we

have that both $\widehat{\sigma}_y^2$ and $\widehat{\gamma}_l$ are quantities that do not depend on ϕ^*, implying that $\widehat{\sigma}_y^2 \xrightarrow{p} \sigma_y^2$ and $\widehat{\gamma}_l \xrightarrow{p} \gamma_l$ for all $\phi^* \in \mathbb{B}$ as $T \longrightarrow \infty$. These are the population moments generated by the ARMA(1,1) model and therefore depend only on β and σ_u^2. Second, we have that $\widehat{\zeta}_u^2$, $\widehat{\delta}_0$, $\widehat{\delta}_l$ and $\widehat{\xi}_l$ for $l \geq 1$ converge to finite quantities. Note that we do not require these quantities to converge to moments evaluated at the true vector of parameters β, but to some finite quantities that will depend on ϕ^*. Hence, considering some vector of estimates ϕ^*, we have that as $T \longrightarrow \infty$, the weak law of large numbers yields:

$$\widehat{\delta}_0 \xrightarrow{p} \delta_0 = \frac{1}{(1+\beta_1\phi_2^*)}\left[\beta_1\left(\gamma_1 - \phi_1^*\sigma_y^2\right)\sigma_u^2 + \beta_2\left(\gamma_1^* - \left(\phi_1^* + \phi_2^*\right)\sigma_u^2\right)\right] \tag{2.47}$$

$$\widehat{\delta}_l \xrightarrow{p} \delta_l = \beta_1^{l-1}\left[\beta_1\delta_0 + \beta_2\sigma_u^2\right], \quad l \geq 1 \tag{2.48}$$

$$\widehat{\zeta}_u^2 \xrightarrow{p} \zeta_u^2 = \frac{1}{(1-\phi_2^*)}\left[\left(1+\phi_1^{*^2}\right)\sigma_y^2 - 2\phi_1^*\gamma_1 - 2\phi_2^*\delta_1 + 2\phi_1^*\phi_2^*\delta_0\right] \tag{2.49}$$

$$\widehat{\xi}_1 \xrightarrow{p} \xi_1 = \gamma_1 - \phi_1^*\sigma_y^2 - \phi_2^*\delta_0 \tag{2.50}$$

$$\begin{aligned}\widehat{\xi}_l \xrightarrow{p} \xi_l = \gamma_l + \left[\sum_{i=2}^{l}(-1)^{l-2}\ (-1)^{l-1}\left(-\phi_2^*\right)^{l-i}\left(\phi_1^* + \phi_2^*\right)\gamma_{i-1}\right] + \\ + (-1)^l\left[\left(-\phi_2^*\right)^{l-1}\phi_1^*\sigma_y^2 + \left(\phi_2^*\right)^l\delta_0\right], \quad l > 1\end{aligned} \tag{2.51}$$

From Assumption 1, we have that the $\sum_{i=0}^{\infty}|-\phi_2^*| < \infty$, $\sum_{i=0}^{\infty}|-\beta_2| < \infty$ and $\sum_{i=0}^{\infty}|-\beta_1| < \infty$ for all $\phi_2^*, \beta_1, \beta_2 \in \mathbb{B}$, implying that the infinite summations in (2.43), (2.44), (2.45) and (2.46) are finite. Following that, it is enough to show that $\left[\zeta_u^2 - \frac{\left(\delta_0^2\right)^2}{\sigma_y^2}\right]$ is different from zero for all $\phi^*, \beta_1, \beta_2 \in \mathbb{B}$, to obtain $\widehat{V}_T\left(\phi^*\right) = O_p\left(1\right)$ as $T \longrightarrow \infty$. This is equivalent to show that

the solutions of

$$-\left(\phi_1^*+\phi_2^*\right)^2\left(-1+\beta_2\ \left(1-\beta_2+\beta_1\left(-1+\phi_2^*\right)+\phi_2^*\right)\right)]\times \frac{\left[\left(1+\beta_2\left(1+\beta_1+\beta_2+\left(-1+\beta_1\right)\phi_2^*\right)\right)\left(\sigma_u^2\right)^2\right]}{\left[\left(-1+\beta_1^2\right)\left(1+\beta_1\phi_2^*\right)^2\left(-1+\phi_2^{*2}\right)\right]}=0 \tag{2.52}$$

do not satisfy Assumption 1. By solving (2.52), we obtain multiple solutions that depend on the following four parameters: β_1, β_2, ϕ_1^* and ϕ_2^*.

$$\beta_1=\left\{\frac{1-\beta_2+\beta_2^2-\beta_2\phi_2^*}{\beta_2\left(-1+\phi_2^2\right)},\quad \frac{-1-\beta_2-\beta_2^2+\beta_2\phi_2^*}{\beta_2\left(1+\phi_2^2\right)}\right\} \tag{2.53}$$

$$\beta_2=\left\{\begin{array}{c}\frac{1}{2}\left[1-\beta_1+\phi_2^*+\beta_1\phi_2^*\pm\sqrt{-4+\left(-1+\beta_1-\phi_2^*-\beta_1\phi_2^*\right)^2}\right],\\ \frac{1}{2}\left[-1-\beta_1+\phi_2^*-\beta_1\phi_2^*\pm\sqrt{-4+\left(1+\beta_1-\phi_2^*+\beta_1\phi_2^*\right)^2}\right]\end{array}\right\} \tag{2.54}$$

$$\phi_2^*=\left\{\frac{-1-\beta_2-\beta_2\beta_1-\beta_2^2}{(-1+\beta_1)\beta_2},\quad \frac{1-\beta_2+\beta_2\beta_1+\beta_2^2}{(1+\beta_1)\beta_2}\right\} \tag{2.55}$$

$$\phi_1^*=-\phi_2^* \tag{2.56}$$

Solution (2.56) is ruled out by Assumption 1. We tackle the remaining solutions through a numerical grid search. We show that there are no real numbers satisfying both Assumption 1 and the set of solutions given by (2.53), (2.54), (2.55), and (2.56).[6] This implies that $\widehat{V}_T\left(\phi^*\right)=O_p\left(1\right)$, yielding that the $\left\|\widehat{V}_T\left(\phi^*\right)\right\|=O_p\left(1\right)$, which proves Lemma 2.

Lemma 3 *Suppose Assumptions 1 and 2 hold. Then,*

$\sup_{\phi\in\mathbb{B}}\left|\widehat{N}_T\left(\phi\right)-N\left(\phi\right)\right|=o_p\left(1\right)$ *as* $T\longrightarrow\infty$

Proof of Lemma 3: We start the proof by showing that the population and sample mapping converge point-wise in probability for all $\phi\in\mathbb{B}$. Fix-

[6] As a matter of space, we do not report the graphs containing the different combinations of parameters satisfying Assumption 1 and the corresponding results of (2.53), (2.54) and (2.55). These results are available upon request.

ing $\beta^{j(T)} \in \mathbb{B}$, as the vector of estimates obtained at the j^{th} iteration, and provided that $T \longrightarrow \infty$, we have:

$$\left| N\left(\beta^{j(T)}\right) - \widehat{N}_T\left(\beta^{j(T)}\right) \right| = \left| \left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \left[\frac{X_{-1}^{j\prime} Y}{T} \right] - \mathbb{E}\left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \mathbb{E}\left[\frac{X_{-1}^{j\prime} Y}{T} \right] \right| \tag{2.57}$$

Defining $\widetilde{N}_T\left(\beta^{j(T)}\right) = \left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \mathbb{E}\left[\frac{X_{-1}^{j\prime} Y}{T} \right]$, we bound (2.57) as:

$$\left| N\left(\beta^{j(T)}\right) - \widehat{N}_T\left(\beta^{j(T)}\right) \right| \leq \left| \left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \left[\frac{X_{-1}^{j\prime} Y}{T} - \mathbb{E}\left[\frac{X_{-1}^{j\prime} Y}{T} \right] \right] \right| + \left| \left[\left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} - \mathbb{E}\left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \right] \times \mathbb{E}\left[\frac{X_{-1}^{j\prime} Y}{T} \right] \right| \tag{2.58}$$

We need to show that both terms on the right-hand side of (2.58) converge in probability to zero. Provided that $T \longrightarrow \infty$, this task becomes a law-of-large-numbers problem, where it suffices to show that:

$$\left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right]^{-1} \xrightarrow{p} \mathbb{E}\left[\frac{X_{-1}^{j\prime} X_{-1}^{j}}{T} \right] \tag{2.59}$$

$$\left[\frac{X_{-1}^{j\prime} Y}{T} \right] \xrightarrow{p} \mathbb{E}\left[\frac{X_{-1}^{j\prime} Y}{T} \right] \tag{2.60}$$

Assumptions 1 and 2 guarantee that the ARMA(1,1) model is covariance-stationary. This allows us to use the weak law of large numbers, such that (2.59) and (2.60) hold for each $\phi \in \mathbb{B}$ as $T \longrightarrow \infty$. Lemma 3, however, requires uniform convergence in probability. To this purpose, the sample mapping needs to be continuous and stochastically equicontinuous for all

$\phi \in \mathbb{B}$. Evoking Lemma 2, we are in a position to apply theorem 21.9 (pg. 337) in Davidson (1994), yielding the final result of this Lemma:

$$\sup_{\phi \in \mathbb{B}} \left| \widehat{N}_T(\phi) - N(\phi) \right| \xrightarrow{p} 0 \tag{2.61}$$

Lemma 4 *Suppose Assumptions 1 and 2 hold. Then,*

$\sup_{\phi,\gamma \in \mathbb{B}} \left| \widehat{\Lambda}_T(\phi,\gamma) - \Lambda(\phi,\gamma) \right| = o_p(1)$ *as* $T \longrightarrow \infty$

Proof of Lemma 4: Fix $\phi, \gamma \in \mathbb{B}$ and rewriting the difference between the population and sample mappings, evaluated at different vector of estimates, using the mean value theorem, we have:

$$\sup_{\phi,\gamma \in \mathbb{B}} |N(\phi) - N(\gamma)| = \sup_{\phi,\gamma \in \mathbb{B}} |\Lambda(\phi,\gamma)[\phi - \gamma]| \tag{2.62}$$

$$\sup_{\phi,\gamma \in \mathbb{B}} \left| \widehat{N}_T(\phi) - \widehat{N}_T(\gamma) \right| = \sup_{\phi,\gamma \in \mathbb{B}} \left| \widehat{\Lambda}_T(\phi,\gamma)[\phi - \gamma] \right| \tag{2.63}$$

with $\Lambda(\phi,\gamma) = \int_0^1 V(\phi + \xi(\phi - \gamma))\, d\xi$ and its sample counterpart defined as $\widehat{\Lambda}_T(\phi,\gamma) = \int_0^1 \widehat{V}_T(\phi + \xi(\phi - \gamma))\, d\xi$. Subtracting (2.63) from (2.62) we have:

$$\begin{aligned}
\sup_{\phi,\gamma \in \mathbb{B}} \left| \Lambda(\phi,\gamma) - \widehat{\Lambda}_T(\phi,\gamma) \right| |\phi - \gamma| \leq & \sup_{\phi,\gamma \in \mathbb{B}} \left| N(\phi) - \widehat{N}_T(\phi) \right| + \\
& \sup_{\phi,\gamma \in \mathbb{B}} \left| N(\gamma) - \widehat{N}_T(\gamma) \right| \\
\sup_{\phi,\gamma \in \mathbb{B}} \left| \Lambda(\phi,\gamma) - \widehat{\Lambda}_T(\phi,\gamma) \right| \leq & \frac{1}{|\phi - \gamma|} \left[\sup_{\phi,\gamma \in \mathbb{B}} \left| N(\phi) - \widehat{N}_T(\phi) \right| + \right. \\
& \left. \sup_{\phi,\gamma \in \mathbb{B}} \left| N(\gamma) - \widehat{N}_T(\gamma) \right| \right]
\end{aligned} \tag{2.64}$$

From Lemma 3, we have that both terms inside the brackets in (2.64) have order $o_p(1)$. Provided that $[\phi - \gamma]$ is bounded and $T \longrightarrow \infty$, we have that

$\sup_{\phi,\gamma\in\mathbb{B}}\left|\Lambda(\phi,\gamma)-\widehat{\Lambda}(\phi,\gamma)\right|=o_p(1)$, proving Lemma 4.

Lemma 5 *Suppose Assumptions 1 and 2 hold and* $\left|\frac{\beta_1\beta_2}{1+\beta_1\beta_2}\right|<1$. *If*

i. $\sup_{\phi\in\mathbb{B}}\left|\widehat{N}_T(\phi)-N(\phi)\right|=o_p(1)$ *as* $T\longrightarrow\infty$

ii. $\sup_{\phi,\gamma\in\mathbb{B}}\left|\widehat{\Lambda}_T(\phi,\gamma)-\Lambda(\phi,\gamma)\right|=o_p(1)$ *as* $T\longrightarrow\infty$

then, $\widehat{N}_T(\phi)$ *is an ACM on* $(\mathbb{B},d)$*, with* $\phi\in\mathbb{B}$ *and it has fixed point denoted by* $\widehat{\beta}$*, such that* $\left|\widehat{\beta}^{j(T)}-\widehat{\beta}\right|=o_p(1)$ *as* $j(T)\longrightarrow\infty$ *with* $T\longrightarrow\infty$.

Proof of Lemma 5: Provided that $N(\phi)$ is an ACM on $(\mathbb{B},d)$, with $\phi\in\mathbb{B}$, we have that $|N(\phi)-N(\gamma)|\leq\kappa|\phi-\gamma|$ holds for each $\phi,\gamma\in\mathbb{B}$. Following that, we bound $\left|\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right|$ as:

$$\left|\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right|\leq|N(\phi)-N(\gamma)|+\left|\left[\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right]-[N(\phi)-N(\gamma)]\right| \tag{2.65}$$

$$\left|\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right|\leq\kappa|\phi-\gamma|+\left|\left[\widehat{\Lambda}_T(\phi,\gamma)-\Lambda(\phi,\gamma)\right][\phi-\gamma]\right| \tag{2.66}$$

From Lemma 4, the second term on the right-hand of equation (2.66) has order $o_p(1)$. Thus as $T\longrightarrow\infty$, we have that $\left|\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right|\leq\kappa|\phi-\gamma|$ yielding the first result of Lemma 4. The second step of the proof consists of showing that $\widehat{\beta}^{j(T)}$ converges to the fixed point $\widehat{\beta}$ as $j(T)\longrightarrow\infty$ with $T\longrightarrow\infty$. To this purpose, we rewrite $\left|\widehat{\beta}^{j(T)}-\widehat{\beta}\right|$ using two implications from the ACM properties of $\widehat{N}_T(\phi)$: $\widehat{N}_T(\phi)$ has a fixed point, such that $\widehat{\beta}=\widehat{N}_T\left(\widehat{\beta}\right)$, and $\left|\widehat{N}_T(\phi)-\widehat{N}_T(\gamma)\right|\leq\widehat{\kappa}|\phi-\gamma|$ holds for each $\phi,\gamma\in\mathbb{B}$ with $|\widehat{\kappa}|\in[0,1)$.

$$\left|\widehat{\beta}^{j(T)}-\widehat{\beta}\right|=\left|\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right)-\widehat{N}_T\left(\widehat{\beta}\right)\right|\leq\widehat{\kappa}\left|\widehat{\beta}^{j(T)-1}-\widehat{\beta}\right| \tag{2.67}$$

Substituting recursively and using the ACM property, we have

$$\left|\widehat{\beta}^{j(T)} - \widehat{\beta}\right| \leq \widehat{\kappa}^{j(T)} \left|\widehat{\beta}^{0} - \widehat{\beta}\right| \tag{2.68}$$

The proof is complete in (2.68), provided that $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$.

Lemma 6 *Suppose Assumptions 1 and 2 hold and* $\left|\frac{\beta_1\beta_2}{1+\beta_1\beta_2}\right| < 1$. *If*

i. $\widehat{N}_T(\phi)$ *is an ACM on* $(\mathbb{B}, d)$

Then, $\sqrt{T}\left|\widehat{\beta}^{j(T)} - \widehat{\beta}\right| = o_p(1)$ *as* $T \longrightarrow \infty$ *and* $j(T) \longrightarrow \infty$

Proof of Lemma 6: We show the $\sqrt{T}$ convergence of $\widehat{\beta}^{j(T)}$ to the fixed point $\widehat{\beta}$ by using the result that yields that the sample mapping is an ACM on $(\mathbb{B}, d)$ and similar steps as in item (i) in Theorem 1. Denote $\widehat{\kappa}$ as the sample counterpart of κ. Then,

$$\begin{aligned}
&\sqrt{T}\left|\widehat{\beta}^{j(T)} - \widehat{\beta}\right| = \sqrt{T}\left|\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right) - \widehat{N}_T\left(\widehat{\beta}\right)\right| \\
&\sqrt{T}\left|\widehat{N}_T\left(\widehat{\beta}^{j(T)-1}\right) - \widehat{N}_T\left(\widehat{\beta}\right)\right| \leq \sqrt{T}\left[\hat{\kappa}\left|\widehat{\beta}^{j(T)-1} - \widehat{\beta}\right|\right]
\end{aligned} \tag{2.69}$$

Substituting recursively (2.69), we have

$$\sqrt{T}\left|\widehat{\beta}^{j(T)} - \widehat{\beta}\right| \leq \sqrt{T}\left[\widehat{\kappa}^{j(T)}\left|\widehat{\beta}^{0} - \widehat{\beta}\right|\right] \tag{2.70}$$

To make the right-hand side of (2.70) converge in probability to zero, we require that $\hat{\kappa}^{j(T)}$ dominates $\sqrt{T}$ as $j(T) \longrightarrow \infty$ with $T \longrightarrow \infty$. A sufficient rate implying this dominance is one such that $j \gg -\frac{1}{2}\left[\frac{\ln(T)}{\ln(\kappa)}\right]$. Hence, provided that $\frac{\ln(T)}{j} = o(1)$, we have that $\sqrt{T}\left|\widehat{\beta}^{j(T)} - \widehat{\beta}\right| = o_p(1)$, which

proves the lemma.

Proof of Theorem 1: We start proving the consistency of the IOLS estimator. From Dominitz and Sherman (2005), if $N(\phi)$ is an ACM on $(\mathbb{B}, d)$, then $N(\phi)$ is also a contraction map. We prove item (i) in Theorem 1 by using the standard fixed-point theorem as stated in Burden and Faires (1997) and Judd (1998). From Lemma 1, $N(\phi)$ is an ACM implying that $\left|N\left(\beta^{j(T)-1}\right) - N(\beta)\right| \leq \kappa \left|\beta^{j(T)-1} - \beta\right|$ holds. Identification on the population mapping gives $N(\beta) = \beta$. Lemma 5 yields that the sample counterpart of $N(\phi)$ is also an ACM on $(\mathbb{B}, d)$ with a fixed point $\widehat{\beta}$.

$$\left|\widehat{\beta} - \beta\right| \leq \left|\beta^{j(T)} - \beta\right| + \left|\widehat{\beta} - \beta^{j(T)}\right| \tag{2.71}$$

We first show that the first term on the right-hand side of (2.71) converges in probability to zero. To this purpose, we rewrite $\left|\beta^{j(T)} - \beta\right|$ as in (2.72), provided that $N(\phi)$ is an ACM and thus $N(\beta) = \beta$.

$$\left|\beta^{j(T)} - \beta\right| = \left|N\left(\beta^{j(T)-1}\right) - N(\beta)\right| \leq \kappa \left|\beta^{j(T)-1} - \beta\right| \tag{2.72}$$

Substituting recursively equation (2.72), the fixed-point theorem result states that as the number of iterations tends to infinity, the sequence converges to the fixed point.

$$\left|\beta^{j(T)} - \beta\right| \leq \kappa^{j(T)} \left|\beta^{0} - \beta\right| \tag{2.73}$$

Thus, provided that $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$, (2.73) yields that $\left|\beta^{j(T)} - \beta\right| = o_p(1)$, and therefore that the first term on the right-hand side of (2.71)

converges in probability to zero.

We now turn our attention to the second term on the right-hand side of (2.71). It remains to show that this term has also order $o_p(1)$. We bound this term using result in Lemma 5.

$$\begin{aligned}\left|\widehat{\beta}-\beta^{j(T)}\right| &= \left|\widehat{N}_T\left(\widehat{\beta}\right)-N\left(\beta^{j(T)-1}\right)\right| \\ &\leq \left|\widehat{N}_T\left(\widehat{\beta}\right)-N\left(\widehat{\beta}\right)\right|+\left|N\left(\widehat{\beta}\right)-N\left(\beta^{j(T)-1}\right)\right|\end{aligned} \tag{2.74}$$

If Lemma 3 holds, then $\left|\widehat{N}_T\left(\widehat{\beta}\right)-N\left(\widehat{\beta}\right)\right| \leq \sup_{\phi\in\mathbb{B}}\left|N(\phi)-\widehat{N}_T(\phi)\right|$ for each $\phi \in \mathbb{B}$, implying that (2.74) resumes to:

$$\left|\widehat{\beta}-\beta^{j(T)}\right| \leq \sup_{\phi\in\mathbb{B}}\left|N(\phi)-\widehat{N}_T(\phi)\right|+\kappa\left|\widehat{\beta}-\beta^{j(T)-1}\right| \tag{2.75}$$

Applying the same steps as in (2.74) and provided that $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$ and $\kappa \in (0,1\,]$

$$\left|\widehat{\beta}-\beta^{j(T)}\right| \leq \sup_{\phi\in\mathbb{B}}\left|N(\phi)-\widehat{N}_T(\phi)\right|\left[1+\kappa+\kappa^2+...\right] \tag{2.76}$$

$$\left|\widehat{\beta}-\beta^{j(T)}\right| \leq \sup_{\phi\in\mathbb{B}}\left|N(\phi)-\widehat{N}_T(\phi)\right|\left[\frac{1}{1-\kappa}\right] \tag{2.77}$$

Because the second term in brackets on the right-hand side of (2.77) is bounded and Lemma 3 yields that the first term has order $o_p(1)$, we have that the fixed point from the sample mapping is a consistent estimate of β, provided that $j(T) \longrightarrow \infty$ as $T \longrightarrow \infty$.

We now turn our attention to show the asymptotic distribution of the IOLS estimator. This proof mirrors the steps of Theorem 4 in Dominitz and Sherman (2005). Similar steps are found in theorem 3.1 in Newey and McFadden (1994). We are interested in establishing the asymptotic

distribution of $\sqrt{T}\left[\widehat{\beta}-\beta\right]$. To this purpose, we write

$$\sqrt{T}\left[\widehat{\beta}^{j(T)}-\beta\right]=\sqrt{T}\left[\widehat{\beta}^{j(T)}-\widehat{\beta}\right]+\sqrt{T}\left[\widehat{\beta}-\beta\right] \tag{2.78}$$

The first term on the right-hand side of equation (2.78) has order $o_p(1)$ following Lemma 6 and provided that $\frac{\ln(T)}{j}=o(1)$. Rewriting the remaining term as

$$\begin{aligned}\sqrt{T}\left[\widehat{\beta}-\beta\right]=&\sqrt{T}\left[\widehat{N}_T\left(\widehat{\beta}\right)-N(\beta)\right]\\ \sqrt{T}\left[\widehat{N}_T\left(\widehat{\beta}\right)-N(\beta)\right]=&\sqrt{T}\left[\left[\widehat{N}_T\left(\widehat{\beta}\right)-\widehat{N}_T(\beta)\right]+\left[\widehat{N}_T(\beta)-\beta\right]\right]\end{aligned} \tag{2.79}$$

Rewriting the first term on the right-hand side of (2.79) into:

$$\left[\widehat{N}_T\left(\widehat{\beta}\right)-\widehat{N}_T(\beta)\right]=\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)\left[\widehat{\beta}-\beta\right] \tag{2.80}$$

such that $\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)=\int_0^1\widehat{V}_T\left(\widehat{\beta}+\xi\left(\widehat{\beta}-\beta\right)\right)d\xi$. Substituting it back into (2.79) and rearranging terms, we have:

$$\begin{aligned}\sqrt{T}\left[\widehat{\beta}-\beta\right]=&\sqrt{T}\left[\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)\left[\widehat{\beta}-\beta\right]\right]+\sqrt{T}\left[\widehat{N}_T(\beta)-\beta\right]\\ \sqrt{T}\left[\widehat{\beta}-\beta\right]=&\sqrt{T}\left[\left[I_2-\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)\right]^{-1}\left[\widehat{N}_T(\beta)-\beta\right]\right]\end{aligned} \tag{2.81}$$

As in Dominitz and Sherman (2005), we first show that $\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)\xrightarrow{p}V(\beta)$. To this purpose, we write $\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)$ as:

$$\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)=V(\beta)+\left[\Lambda\left(\widehat{\beta},\beta\right)-V(\beta)\right]+\left[\widehat{\Lambda}_T\left(\widehat{\beta},\beta\right)-\Lambda\left(\widehat{\beta},\beta\right)\right] \tag{2.82}$$

From item (i) in Theorem 1 we have that $\widehat{\beta}$ converges in probability to β

as $j \longrightarrow \infty$ with $T \longrightarrow \infty$. This implies that $\Lambda\left(\widehat{\beta}, \beta\right) \xrightarrow{p} V(\beta)$, yielding that the second term on the right-hand converges in probability to zero. Lemma 4 implies that the last term on the right-hand side of (2.82) has order $o_p(1)$, providing that $\widehat{\Lambda}_T\left(\widehat{\beta}, \beta\right) \xrightarrow{p} V(\beta)$. From this result, (2.81) reduces to:

$$\sqrt{T}\left[\widehat{\beta} - \beta\right] = \sqrt{T}\left[\left[I_2 - V(\beta)\right]^{-1}\left[\widehat{N}_T(\beta) - \beta\right]\right] \tag{2.83}$$

Hence, as $T \longrightarrow \infty$ it remains to study the asymptotic distribution of $\sqrt{T}\left[\widehat{N}_T(\beta) - \beta\right]$. To this purpose, we write:

$$\begin{aligned}
\sqrt{T}\left[\widehat{N}_T(\beta) - \beta\right] &= \sqrt{T}\left[\left[X'_{-1}X_{-1}\right]^{-1} X'_{-1}U\right] \\
\sqrt{T}\left[\left[X'_{-1}X_{-1}\right]^{-1} X'_{-1}U\right] &= \left[\left[\frac{X'_{-1}X_{-1}}{T}\right]^{-1}\left[\frac{1}{\sqrt{T}}\right] X'_{-1}U\right]
\end{aligned} \tag{2.84}$$

Applying the Central Limit Theorem for martingale difference sequences and provided that plim $\left[\frac{X'_{-1}X_{-1}}{T}\right]^{-1} = H^{-1}$, it follows that

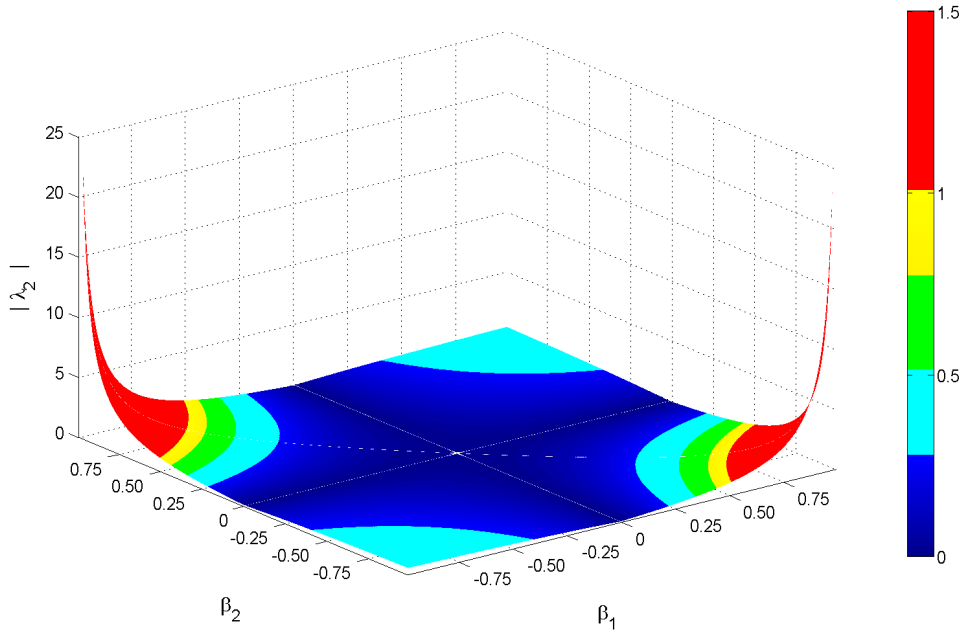
$$\sqrt{T}\left[\widehat{N}_T(\beta) - \beta\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2 H^{-1}\right) \tag{2.85}$$

We conclude the proof of Theorem 1 by setting $A = \left[I - V(\beta)\right]^{-1}$ in (2.83) and using the result in (2.85), such that:

$$\sqrt{T}\left[\widehat{\beta} - \beta\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2 A H^{-1} A'\right) \tag{2.86}$$

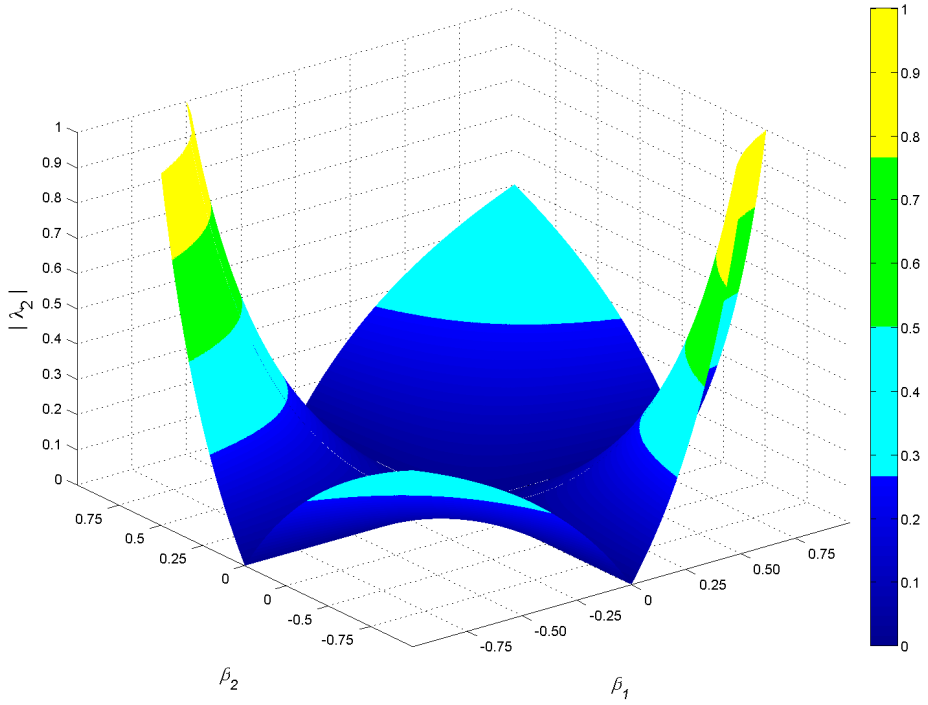
■

Figure 2.1: Maximum Eigenvalues of $V(\beta)$



We plot $|\lambda_2|$ computed using different combinations of β_1 and β_2 such that Assumption 1 is satisfied. For viewing purposes, we truncate the parameter interval in this analysis such that $\beta_1 = [-0.980, 0.980]$ and $\beta_2 = [-0.980, 0.980]$. The grid is fixed in 0.001.

Figure 2.2: Maximum Eigenvalues of $V(\beta)$ - trimmed version



We plot $|\lambda_2|$ computed using different combinations of β_1 and β_2 such that Assumption 1 is satisfied and $|\lambda_2| < 1$. The grid is fixed in 0.001.

Table 2.1: Monte Carlo - Consistency and Efficiency: Small Datasets

K=3															
$eig(A) = eig(M) = (0.5, ..., 0.5)'$															
		T = 50			T = 100			T = 150			T = 200			T = 400		
	True Values	IOLS	MLE	RelMSE	IOLS	MLE	RelMSE	IOLS	MLE	RelMSE	IOLS	MLE	RelMSE	IOLS	MLE	RelMSE
A_{11}	1.00	0.99	0.98	0.84	0.99	0.99	0.58	0.99	0.99	1.01	0.99	0.99	1.03	0.99	0.99	1.03
A_{12}	0.23	0.25	0.28	0.53	0.24	0.25	0.79	0.24	0.25	0.92	0.24	0.24	0.95	0.24	0.24	0.96
A_{13}	-0.05	-0.05	-0.05	0.90	-0.05	-0.05	0.44	-0.05	-0.05	1.00	-0.05	-0.05	1.01	-0.05	-0.05	1.04
A_{21}	-0.91	-0.93	-0.91	1.13	-0.92	-0.90	0.11	-0.91	-0.91	1.49	-0.91	-0.91	1.55	-0.91	-0.91	1.57
A_{22}	0.07	0.12	0.07	0.73	0.09	0.06	0.12	0.08	0.07	1.02	0.08	0.07	1.02	0.08	0.07	1.04
A_{23}	0.09	0.07	0.07	0.96	0.08	0.08	0.56	0.08	0.08	1.19	0.08	0.08	1.24	0.08	0.09	1.32
A_{31}	0.75	0.77	0.74	0.69	0.77	0.76	0.74	0.76	0.75	1.19	0.76	0.75	1.29	0.75	0.75	1.31
A_{32}	0.35	0.27	0.44	0.34	0.32	0.37	0.86	0.33	0.36	0.94	0.34	0.36	0.99	0.34	0.36	1.01
A_{33}	0.43	0.38	0.40	0.87	0.39	0.39	0.90	0.40	0.40	1.09	0.41	0.41	1.17	0.42	0.42	1.22
M_{11}	0.53	0.35	0.37	1.05	0.42	0.43	0.35	0.45	0.46	1.18	0.46	0.47	1.21	0.49	0.49	1.25
M_{12}	-0.95	-0.71	-0.75	1.13	-0.79	-0.81	0.92	-0.83	-0.84	1.12	-0.85	-0.86	1.11	-0.89	-0.89	1.11
M_{13}	-0.50	-0.33	-0.35	1.11	-0.39	-0.41	0.78	-0.41	-0.42	1.17	-0.43	-0.44	1.17	-0.46	-0.46	1.16
M_{21}	-0.01	0.05	0.02	0.92	0.03	0.01	1.26	0.02	0.01	1.65	0.01	0.01	1.73	0.01	0.00	1.90
M_{22}	0.86	0.54	0.68	1.28	0.65	0.78	0.11	0.70	0.78	1.74	0.73	0.79	1.68	0.78	0.81	1.57
M_{23}	0.19	0.11	0.14	0.79	0.13	0.15	0.37	0.14	0.15	1.09	0.14	0.15	1.13	0.16	0.16	1.25
M_{31}	0.02	0.03	0.05	0.70	0.03	0.03	0.42	0.03	0.03	1.52	0.03	0.03	1.67	0.03	0.03	1.81
M_{32}	-0.74	-0.76	-0.93	0.31	-0.79	-0.83	0.60	-0.79	-0.81	0.91	-0.78	-0.79	0.97	-0.77	-0.77	1.04
M_{33}	0.11	0.06	0.05	0.70	0.09	0.10	0.69	0.09	0.10	1.11	0.10	0.10	1.20	0.10	0.10	1.31
RT		282			677			477			576			239		
%F		38%	42%		17%	5%		11%	2%		8%	1%		3%	0%	

RT is Relative time of computation (MLE/IOLS); %F is the percentage of non-convergence replications. For the IOLS algorithm, failure happens when the algorithm does not converge. For the MLE estimator, we assume failure happens when either the return code is different than zero or estimates do not satisfy the covariance stationarity condition; MLE estimation is performed without any constraint. We report the mean and RelMSE computed within all replications that converged. We compute the RelMSE as the ratio of MSE measures obtained from the IOLS and MLE estimators. We denote $A_{i,j}$ as the element in the i^{th} row and j^{th} column of matrix A. All Kronecker indices are set equal to one. When $T = 50$, we perform 3000 replications. For all the remaining scenarios the number of replications is set to be equal to 5000.

Table 2.2: Monte Carlo - Consistency and Efficiency: Medium Datasets

	K = 8					K = 10				
T	MC repl	Compt. Days	% F IOLS	% F MLE	RT	MC repl	Compt. Days	% F IOLS	% F MLE	RT
100	200	14.1	67.5%	99.5%	-	100	12.0	96.0%	100.0%	-
150	200	22.8	48.5%	88.5%	76.7	100	11.3	81.0%	100.0%	-
200	200	25.1	28.5%	40.0%	175.0	78	15.0	55.0%	100.0%	-
400	120	22.6	6.3%	0.0%	469.8	15	27.7	26.7%	53.3%	17,336.6

RT is Relative time of computation (MLE/MLE); %F is percentage of non-convergence replications. For the IOLS algorithm, failure happens when the algorithm does not converge. For the MLE estimator, we assume failure happens when either the return code is different than zero or estimates do not satisfy the covariance stationarity condition; MLE estimation is performed without any constraint. Compt. Days accounts for the time of computation measured in days and MC repl. accounts for the number of replications.

Table 2.3: Monte Carlo - IOLS Consistency: Medium Datasets with Low Eigenvalues

K = 10 $eig(A) = eig(M) = (0.3, ..., 0.3)'$																	
	T = 150					T = 200				T = 400				T = 1000			
		IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c
$A_{1,1}$	0.054	0.146	0.145	0.102	0.182	0.072	0.107	0.064	0.110	0.078	0.118	0.068	0.122	0.064	0.078	0.061	0.078
$A_{3,3}$	0.436	0.552	0.166	0.470	0.177	0.445	0.099	0.438	0.100	0.450	0.110	0.441	0.112	0.440	0.072	0.437	0.072
$A_{5,5}$	0.275	0.331	0.126	0.288	0.166	0.275	0.089	0.270	0.091	0.276	0.099	0.270	0.101	0.273	0.063	0.271	0.064
$A_{7,7}$	0.376	0.382	0.110	0.365	0.172	0.371	0.098	0.367	0.101	0.369	0.109	0.363	0.114	0.371	0.071	0.370	0.072
$A_{9,9}$	0.127	0.205	0.139	0.159	0.184	0.134	0.112	0.126	0.115	0.136	0.123	0.128	0.128	0.127	0.081	0.124	0.081
$A_{10,10}$	0.391	0.411	0.140	0.398	0.228	0.394	0.133	0.389	0.137	0.395	0.147	0.388	0.154	0.389	0.098	0.386	0.099
$M_{1,1}$	0.260	-0.008	0.371	0.157	0.217	0.225	0.118	0.232	0.118	0.215	0.131	0.223	0.132	0.240	0.083	0.242	0.083
$M_{3,3}$	0.425	0.119	0.401	0.283	0.233	0.370	0.115	0.371	0.116	0.359	0.130	0.361	0.132	0.391	0.080	0.392	0.080
$M_{5,5}$	0.173	-0.065	0.347	0.086	0.203	0.138	0.102	0.142	0.102	0.131	0.116	0.135	0.115	0.152	0.071	0.153	0.071
$M_{7,7}$	0.119	-0.059	0.307	0.061	0.200	0.095	0.108	0.098	0.110	0.092	0.121	0.095	0.124	0.106	0.077	0.106	0.078
$M_{9,9}$	0.207	-0.042	0.359	0.125	0.214	0.179	0.120	0.187	0.121	0.172	0.134	0.180	0.136	0.195	0.084	0.197	0.085
$M_{10,10}$	0.382	0.154	0.345	0.218	0.289	0.319	0.153	0.313	0.159	0.309	0.171	0.302	0.180	0.345	0.109	0.344	0.111
%F	80%					23%				27%				12%			

We report the mean and RMSE computed within all replications. IOLS^c and RMSE^c denote the measures computed using only the replications that converged. We denote $A_{i,j}$ as the element in the i^{th} row and j^{th} column of matrix A. All Kronecker indices are set equal to one and models are estimated with the IOLS algorithm. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000.

Table 2.4: Monte Carlo - IOLS Consistency: Medium Datasets with High Eigenvalues

K = 10 $eig(A) = eig(M) = (0.8, ..., 0.8)'$																	
		T = 150				T = 200				T = 400				T = 1000			
		IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c
$A_{1,1}$	0.755	0.805	0.075	0.709	0.081	0.811	0.076	0.714	0.072	0.742	0.039	0.733	0.043	0.746	0.025	0.745	0.025
$A_{3,3}$	0.775	0.818	0.069	0.733	0.083	0.824	0.070	0.736	0.067	0.761	0.038	0.754	0.041	0.767	0.023	0.766	0.023
$A_{5,5}$	0.770	0.811	0.067	0.722	0.084	0.819	0.067	0.738	0.063	0.758	0.038	0.750	0.041	0.762	0.023	0.762	0.023
$A_{7,7}$	0.792	0.831	0.068	0.738	0.084	0.839	0.068	0.749	0.071	0.776	0.040	0.768	0.043	0.782	0.024	0.782	0.024
$A_{9,9}$	0.848	0.879	0.061	0.796	0.083	0.886	0.060	0.810	0.065	0.833	0.038	0.827	0.041	0.840	0.022	0.839	0.022
$A_{10,10}$	0.841	0.878	0.063	0.802	0.072	0.883	0.063	0.806	0.063	0.828	0.036	0.822	0.039	0.833	0.021	0.832	0.021
$M_{1,1}$	0.755	0.109	0.716	0.626	0.164	0.233	0.584	0.655	0.131	0.696	0.083	0.701	0.079	0.727	0.045	0.728	0.045
$M_{3,3}$	0.847	0.196	0.719	0.652	0.220	0.325	0.582	0.695	0.175	0.758	0.108	0.753	0.114	0.798	0.065	0.796	0.067
$M_{5,5}$	0.794	0.152	0.713	0.651	0.173	0.283	0.572	0.684	0.137	0.733	0.085	0.734	0.084	0.767	0.045	0.767	0.045
$M_{7,7}$	0.895	0.226	0.734	0.663	0.253	0.363	0.585	0.707	0.206	0.763	0.145	0.771	0.137	0.818	0.086	0.819	0.086
$M_{9,9}$	0.781	0.139	0.713	0.613	0.197	0.264	0.576	0.648	0.161	0.700	0.101	0.700	0.102	0.736	0.059	0.736	0.059
$M_{10,10}$	0.777	0.132	0.712	0.603	0.205	0.268	0.569	0.647	0.158	0.710	0.092	0.701	0.101	0.741	0.054	0.740	0.056
%F		97%				93%				60%				14%			

We report the mean and RMSE computed within all replications. IOLS^c and RMSE^c denote the measures computed using only the replications that converged. We denote $A_{i,j}$ as the element in the i^{th} row and j^{th} column of matrix A. All Kronecker indices are set equal to one and models are estimated with the IOLS algorithm. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000. We denote IOLS^c and RMSE^c as the mean and root mean squared error computed using only the replications which achieved convergence.

Table 2.5: Monte Carlo - IOLS Consistency: Medium Datasets with Mixed Eigenvalues

K = 10																	
$eig(A) = eig(M)$ = mixed Eigenvalues																	
		T = 150				T = 200				T = 400				T = 1000			
		IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c
$A_{1,1}$	0.310	0.404	0.143	0.333	0.136	0.390	0.132	0.322	0.112	0.317	0.078	0.310	0.078	0.310	0.047	0.309	0.047
$A_{3,3}$	0.564	0.693	0.171	0.572	0.125	0.668	0.159	0.559	0.108	0.574	0.078	0.560	0.076	0.565	0.048	0.561	0.047
$A_{5,5}$	0.662	0.742	0.118	0.657	0.098	0.727	0.110	0.652	0.083	0.663	0.058	0.654	0.058	0.661	0.036	0.658	0.035
$A_{7,7}$	0.341	0.464	0.170	0.343	0.130	0.442	0.157	0.338	0.111	0.346	0.078	0.334	0.077	0.341	0.048	0.338	0.048
$A_{9,9}$	0.702	0.757	0.094	0.679	0.084	0.748	0.088	0.684	0.070	0.698	0.046	0.690	0.047	0.700	0.028	0.698	0.028
$A_{10,10}$	0.514	0.591	0.112	0.512	0.091	0.577	0.104	0.507	0.075	0.516	0.053	0.508	0.053	0.513	0.032	0.511	0.032
$M_{1,1}$	0.567	0.141	0.495	0.285	0.321	0.222	0.402	0.333	0.267	0.405	0.183	0.408	0.182	0.471	0.111	0.468	0.114
$M_{3,3}$	0.456	0.036	0.513	0.352	0.171	0.152	0.401	0.384	0.136	0.399	0.101	0.415	0.090	0.430	0.056	0.434	0.053
$M_{5,5}$	0.396	0.039	0.450	0.313	0.149	0.136	0.354	0.339	0.118	0.357	0.081	0.366	0.076	0.379	0.047	0.381	0.045
$M_{7,7}$	0.644	0.193	0.530	0.443	0.248	0.296	0.420	0.476	0.206	0.515	0.154	0.528	0.142	0.564	0.095	0.566	0.093
$M_{9,9}$	0.643	0.287	0.442	0.510	0.175	0.376	0.347	0.539	0.145	0.583	0.092	0.579	0.097	0.609	0.056	0.609	0.057
$M_{10,10}$	0.355	-0.012	0.456	0.243	0.164	0.087	0.352	0.270	0.130	0.287	0.097	0.297	0.090	0.315	0.058	0.317	0.057
%F		79%				64%				32%				13%			

We report the mean and RMSE computed within all replications. IOLS^c and RMSE^c denote the measures computed using only the replications that converged.. We denote $A_{i,j}$ as the element in the i^{th} row and j^{th} column of matrix A. All Kronecker indices are set equal to one and models are estimated with the IOLS algorithm. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000. We denote IOLS^c and RMSE^c as the mean and root mean squared error computed using only the replications which achieved convergence.

Table 2.6: Monte Carlo - IOLS Consistency: Large Datasets with Intermediate Eigenvalues

K = 20 $eig(A) = eig(M) = (0.6, ..., 0.6)'$															
	T = 200			T = 300				T = 400				T = 1000			
	True Values	IOLS	RMSE	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c	IOLS	RMSE	IOLS^c	RMSE^c
$A_{1,1}$	0.579	0.678	0.115	0.679	0.121	0.558	0.065	0.636	0.100	0.563	0.056	0.571	0.032	0.570	0.032
$A_{3,3}$	0.603	0.720	0.133	0.720	0.139	0.581	0.071	0.668	0.113	0.585	0.058	0.595	0.035	0.594	0.035
$A_{5,5}$	0.751	0.843	0.107	0.839	0.108	0.732	0.059	0.802	0.089	0.738	0.050	0.744	0.028	0.743	0.028
$A_{7,7}$	0.649	0.741	0.108	0.743	0.114	0.623	0.061	0.702	0.097	0.629	0.055	0.640	0.031	0.639	0.031
$A_{9,9}$	0.667	0.772	0.121	0.774	0.129	0.636	0.071	0.727	0.107	0.646	0.058	0.658	0.033	0.657	0.033
$A_{11,11}$	0.567	0.687	0.134	0.687	0.140	0.546	0.067	0.635	0.115	0.550	0.059	0.559	0.035	0.558	0.035
$A_{13,13}$	0.597	0.696	0.116	0.701	0.126	0.574	0.067	0.653	0.105	0.574	0.060	0.587	0.035	0.586	0.035
$A_{15,15}$	0.603	0.707	0.119	0.708	0.126	0.582	0.063	0.660	0.105	0.582	0.058	0.594	0.033	0.593	0.033
$A_{17,17}$	0.594	0.709	0.130	0.712	0.138	0.576	0.062	0.660	0.114	0.575	0.058	0.585	0.035	0.584	0.035
$A_{19,19}$	0.735	0.804	0.085	0.805	0.086	0.727	0.053	0.774	0.072	0.722	0.042	0.731	0.025	0.731	0.025
$A_{20,20}$	0.600	0.704	0.119	0.705	0.126	0.575	0.066	0.656	0.102	0.580	0.055	0.591	0.032	0.590	0.032
$M_{1,1}$	0.621	0.018	0.638	0.097	0.585	0.486	0.156	0.345	0.336	0.509	0.130	0.550	0.081	0.552	0.079
$M_{3,3}$	0.630	0.038	0.627	0.105	0.588	0.531	0.129	0.362	0.346	0.560	0.094	0.592	0.055	0.593	0.054
$M_{5,5}$	0.589	0.036	0.590	0.101	0.548	0.482	0.135	0.335	0.322	0.505	0.110	0.552	0.057	0.552	0.057
$M_{7,7}$	0.573	0.011	0.600	0.080	0.559	0.509	0.095	0.330	0.328	0.528	0.079	0.556	0.042	0.557	0.041
$M_{9,9}$	0.567	0.000	0.604	0.057	0.579	0.507	0.091	0.324	0.327	0.524	0.077	0.551	0.041	0.552	0.040
$M_{11,11}$	0.579	-0.010	0.624	0.058	0.586	0.502	0.112	0.317	0.344	0.517	0.090	0.548	0.050	0.549	0.049
$M_{13,13}$	0.604	0.013	0.627	0.091	0.576	0.487	0.141	0.336	0.335	0.513	0.112	0.550	0.068	0.552	0.065
$M_{15,15}$	0.567	-0.025	0.627	0.061	0.571	0.484	0.112	0.313	0.334	0.508	0.088	0.533	0.051	0.534	0.050
$M_{17,17}$	0.630	0.039	0.629	0.107	0.588	0.539	0.117	0.366	0.345	0.567	0.089	0.601	0.049	0.602	0.048
$M_{19,19}$	0.547	0.019	0.569	0.077	0.535	0.463	0.117	0.321	0.303	0.496	0.082	0.525	0.044	0.525	0.044
$M_{20,20}$	0.608	0.031	0.614	0.095	0.578	0.525	0.115	0.356	0.331	0.548	0.088	0.583	0.047	0.583	0.047
%F	100%			86%				53%				4%			

We report the mean and RMSE computed within all replications. IOLS^c and RMSE^c denote the measures computed using only the replications that converged. We denote $A_{i,j}$ as the element in the i^{th} row and j^{th} column of matrix A. All Kronecker indices are set equal to one and models are estimated with the IOLS algorithm. We perform 2000 replications in all scenarios. We denote IOLS^c and RMSE^c as the mean and root mean squared error computed using only the replications which achieved convergence. We perform 2000 replications for all sample sizes.

Table 2.7: Monte Carlo - Forecast Exercise: Medium Dataset with Low Eigenvalues

K = 10 $eig(A) = eig(M) = (0.3, ..., 0.3)'$																
	T = 150				T = 200				T = 400				T = 1000			
	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c
H(1)	2.674	2.038	0.562	0.429	0.983	0.371	0.964	0.364	0.982	0.381	0.957	0.371	0.981	0.352	0.978	0.351
H(2)	1.277	1.013	0.541	0.461	0.958	0.420	0.949	0.417	0.954	0.427	0.939	0.422	0.970	0.408	0.969	0.408
H(3)	1.111	1.130	0.544	0.586	0.953	0.543	0.947	0.540	0.947	0.550	0.936	0.545	0.969	0.531	0.967	0.530
H(4)	0.973	1.300	0.544	0.743	0.947	0.697	0.945	0.696	0.939	0.704	0.934	0.700	0.966	0.686	0.965	0.685
H(5)	0.874	1.356	0.556	0.867	0.948	0.825	0.948	0.825	0.940	0.830	0.937	0.828	0.967	0.815	0.966	0.815
H(6)	0.794	1.292	0.576	0.936	0.953	0.903	0.955	0.905	0.946	0.906	0.945	0.905	0.970	0.896	0.970	0.896
H(7)	0.734	1.181	0.601	0.965	0.962	0.943	0.965	0.945	0.956	0.944	0.956	0.943	0.975	0.937	0.976	0.939
H(8)	0.700	1.088	0.627	0.976	0.971	0.961	0.974	0.964	0.967	0.961	0.967	0.961	0.980	0.958	0.982	0.959
H(9)	0.690	1.035	0.652	0.980	0.980	0.971	0.983	0.974	0.977	0.971	0.977	0.971	0.986	0.969	0.988	0.971
H(10)	0.696	1.012	0.676	0.982	0.986	0.977	0.989	0.980	0.984	0.977	0.984	0.977	0.991	0.976	0.993	0.978
H(11)	0.711	1.004	0.698	0.985	0.990	0.982	0.993	0.985	0.989	0.982	0.989	0.983	0.994	0.981	0.997	0.984
H(12)	0.729	1.000	0.720	0.988	0.993	0.986	0.996	0.988	0.992	0.986	0.992	0.986	0.996	0.986	0.999	0.988
%F	80%				23%				27%				12%			

H(h) is the hth-step-ahead forecast. We assess forecast accuracy in terms of the out-of-the-sample RelMSFE. We report the mean of the RelMSFE measures within all variables in the system. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and set all Kronecker indices equal to one. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000.

Table 2.8: Monte Carlo - Forecast Exercise: Medium Dataset with High Eigenvalues

K = 10 $eig(A) = eig(M) = (0.8, ..., 0.8)'$																
	T = 150				T = 200				T = 400				T = 1000			
	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c
H(1)	4.683	8.398	0.426	0.761	1.527	1.527	0.715	0.715	0.914	0.648	0.906	0.642	0.946	0.604	0.946	0.604
H(2)	1.363	3.335	0.377	0.924	0.807	1.106	0.647	0.886	0.874	0.821	0.874	0.821	0.934	0.791	0.934	0.791
H(3)	1.054	2.419	0.398	0.915	0.745	1.010	0.649	0.881	0.872	0.820	0.872	0.820	0.934	0.790	0.934	0.790
H(4)	0.985	2.058	0.424	0.889	0.747	0.965	0.665	0.860	0.878	0.805	0.877	0.804	0.935	0.776	0.936	0.777
H(5)	0.966	1.867	0.447	0.866	0.760	0.939	0.680	0.842	0.882	0.794	0.881	0.793	0.937	0.767	0.938	0.768
H(6)	0.958	1.746	0.466	0.852	0.778	0.926	0.696	0.830	0.890	0.789	0.889	0.788	0.940	0.765	0.941	0.766
H(7)	0.947	1.661	0.482	0.846	0.793	0.920	0.711	0.826	0.898	0.791	0.897	0.789	0.943	0.768	0.943	0.769
H(8)	0.932	1.596	0.494	0.847	0.806	0.920	0.724	0.828	0.906	0.797	0.904	0.796	0.946	0.776	0.947	0.777
H(9)	0.913	1.542	0.504	0.851	0.817	0.923	0.737	0.834	0.913	0.807	0.912	0.806	0.950	0.788	0.951	0.789
H(10)	0.893	1.495	0.513	0.859	0.826	0.930	0.748	0.844	0.920	0.820	0.919	0.818	0.953	0.801	0.954	0.803
H(11)	0.873	1.453	0.521	0.868	0.833	0.938	0.759	0.855	0.927	0.833	0.925	0.832	0.957	0.816	0.958	0.817
H(12)	0.852	1.415	0.529	0.879	0.840	0.946	0.770	0.867	0.932	0.848	0.930	0.846	0.960	0.831	0.961	0.832
%F	97%				93%				60%				14%			

H(h) is the hth-step-ahead forecast. We assess forecast accuracy in terms of the out-of-the-sample RelMSFE. We report the mean of the RelMSFE measures within all variables in the system. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and set all Kronecker indices equal to one. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000.

Table 2.9: Monte Carlo - Forecast Exercise: Medium Dataset with Mixed Eigenvalues

K = 10																
$eig(A) = eig(M)$ = mixed Eigenvalues																
	T = 150				T = 200				T = 400				T = 1000			
	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c
H(1)	2.319	2.737	0.512	0.602	1.696	1.163	0.824	0.567	0.942	0.526	0.931	0.520	0.971	0.497	0.959	0.492
H(2)	0.949	1.206	0.460	0.594	0.990	0.724	0.778	0.574	0.909	0.547	0.908	0.546	0.944	0.531	0.942	0.530
H(3)	0.865	1.054	0.477	0.593	0.940	0.685	0.783	0.575	0.910	0.553	0.910	0.553	0.943	0.540	0.943	0.539
H(4)	0.853	1.074	0.494	0.630	0.938	0.718	0.797	0.614	0.917	0.593	0.917	0.593	0.946	0.579	0.946	0.579
H(5)	0.845	1.128	0.508	0.685	0.939	0.774	0.810	0.670	0.925	0.648	0.926	0.649	0.949	0.634	0.949	0.634
H(6)	0.830	1.171	0.522	0.745	0.939	0.830	0.822	0.729	0.932	0.706	0.933	0.707	0.953	0.691	0.953	0.691
H(7)	0.806	1.186	0.536	0.798	0.935	0.875	0.833	0.783	0.938	0.758	0.940	0.759	0.957	0.744	0.956	0.743
H(8)	0.778	1.177	0.551	0.840	0.928	0.907	0.844	0.828	0.944	0.802	0.946	0.804	0.960	0.788	0.960	0.788
H(9)	0.751	1.151	0.566	0.873	0.922	0.927	0.855	0.862	0.949	0.837	0.951	0.839	0.964	0.825	0.964	0.824
H(10)	0.729	1.120	0.581	0.896	0.918	0.939	0.867	0.887	0.955	0.864	0.957	0.866	0.968	0.853	0.967	0.853
H(11)	0.714	1.088	0.596	0.911	0.917	0.945	0.878	0.906	0.960	0.885	0.962	0.887	0.971	0.875	0.971	0.875
H(12)	0.706	1.061	0.612	0.921	0.918	0.947	0.889	0.918	0.965	0.900	0.967	0.902	0.974	0.891	0.974	0.892
%F	79%				64%				32%				13%			

H(h) is the hth-step-ahead forecast. We assess forecast accuracy in terms of the out-of-the-sample RelMSFE. We report the mean of the RelMSFE measures within all variables in the system. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and set all Kronecker indices equal to one. When $T = 1000$, we perform 2000 replications. For all the remaining scenarios the number of replications is set to be equal to 10000.

Table 2.10: Monte Carlo - Forecast Exercise: Large Dataset with Intermediate Eigenvalues

K = 20 $eig(A) = eig(M) = (0.6, ..., 0.6)'$														
	T = 200		T = 300				T = 400				T = 1000			
	VAR	AR(1)	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c	VAR	AR(1)	VAR^c	AR(1)^c
H(1)	0.476	3.722	4.854	6.557	0.580	0.778	1.230	0.967	0.934	0.734	0.950	0.664	0.950	0.664
H(2)	0.272	1.631	1.651	2.385	0.535	0.774	0.966	0.801	0.898	0.745	0.940	0.698	0.940	0.698
H(3)	0.274	1.256	1.293	1.663	0.552	0.717	0.941	0.730	0.895	0.695	0.939	0.658	0.940	0.658
H(4)	0.271	1.126	1.210	1.439	0.577	0.694	0.951	0.706	0.910	0.676	0.945	0.645	0.945	0.645
H(5)	0.268	1.089	1.190	1.380	0.597	0.704	0.963	0.714	0.925	0.687	0.951	0.658	0.951	0.659
H(6)	0.264	1.095	1.179	1.386	0.613	0.733	0.971	0.741	0.936	0.716	0.957	0.689	0.958	0.689
H(7)	0.259	1.117	1.160	1.410	0.625	0.772	0.975	0.779	0.944	0.755	0.963	0.728	0.963	0.728
H(8)	0.253	1.140	1.131	1.427	0.637	0.814	0.977	0.820	0.949	0.797	0.967	0.770	0.967	0.770
H(9)	0.247	1.154	1.092	1.422	0.648	0.853	0.978	0.858	0.953	0.837	0.970	0.810	0.970	0.810
H(10)	0.239	1.157	1.047	1.395	0.660	0.886	0.979	0.889	0.958	0.871	0.972	0.845	0.973	0.846
H(11)	0.232	1.148	0.999	1.350	0.672	0.912	0.980	0.915	0.963	0.899	0.974	0.875	0.975	0.876
H(12)	0.223	1.132	0.954	1.294	0.684	0.930	0.982	0.934	0.968	0.921	0.977	0.899	0.978	0.900
%F	100%		86%				53%				4%			

H(h) is the hth-step-ahead forecast. We assess forecast accuracy in terms of the out-of-the-sample RelMSFE. We report the mean of the RelMSFE measures within all variables in the system. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and set all Kronecker indices equal to one. We perform 2000 replications for all sample-sizes.

Table 2.11: Datasets Specification

	$K=10$						$K=20$						$K=40$				
DATASET	1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
IPS10	×	×	×	×	×		×	×	×	×	×		×	×	×	×	×
FYFF	×	×	×	×	×		×	×	×	×	×		×	×	×	×	×
PUNEW	×	×	×	×	×		×	×	×	×	×		×	×	×	×	×
A0M052	×				×		×		×	×	×		×	×	×	×	×
A0M051		×					×	×					×	×	×		
A0M224R			×					×		×	×		×	×	×	×	×
A0M057	×							×			×		×		×	×	×
A0M059		×					×		×	×			×	×		×	×
PMP				×			×							×	×	×	
A0m082			×		×			×	×		×		×	×	×		×
LHEL									×		×			×	×	×	×
LHELX			×					×		×			×		×	×	
LHEM								×	×				×	×	×		×
LHUR	×									×	×				×	×	×
CES002				×										×	×		×
A0M048							×						×			×	
PMI									×	×			×	×	×	×	×
PMNO					×									×		×	
PMDEL				×										×	×	×	
PMNV								×					×			×	×
FM1		×					×	×	×				×	×	×	×	×
FM2			×		×			×	×	×			×	×		×	×
FM3				×										×	×		×
FM2DQ	×									×	×		×	×	×	×	×
FMFBA								×	×		×		×	×			×
FMRRA														×	×		
FMRNBA													×		×		×
FCLNQ		×			×		×	×		×			×		×	×	
FCLBMC								×	×		×		×	×	×		×
CCINRV		×					×			×	×			×	×	×	×
A0M095			×				×	×			×		×	×		×	×
FSPCOM			×				×	×			×		×			×	×
FSPIN									×				×	×	×		×
FSDXP										×			×	×		×	×
FSPXE				×									×		×		×
CP90	×						×				×		×	×	×		×
FYGM3	×				×			×		×	×		×	×	×	×	×
FYGM6									×				×	×		×	
FYGT1		×					×			×			×	×		×	×
FYGT5				×					×		×		×	×	×	×	×
FYGT10	×							×					×		×	×	×
FYAAAC		×					×		×	×	×			×	×	×	×
FYBAAC			×					×		×			×	×	×	×	
EXRUS							×		×				×	×	×	×	×
EXRSW										×				×		×	
EXRJAN							×			×			×		×	×	×
EXRUK							×		×				×	×	×	×	×
EXRCAN													×		×	×	
PWFSA				×			×				×			×	×	×	×
PWFCSA					×				×				×	×	×	×	×
PWIMSA													×	×	×	×	×
PWCMSA													×	×	×	×	×

* × indicates that the assigned variable belongs to the specified dataset.

Table 2.12: Forecast: Medium Systems

Hannan-Kavalieris (HM) Algorithm K = 10, T = 470, 50 out of sample forecast							MT Algorithm K = 10, T = 470, 50 out of sample forecast							OZ Algorithm K = 10, T = 470, 50 out of sample forecast						
	Hor:1			Hor:4				Hor:1			Hor:4				Hor:1			Hor:4		
dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.079	1.147	1.120	1.028	1.117*	1.075	IPS10	0.764**	0.812*	0.885**	0.846**	0.920**	0.793**	IPS10	1.031	1.095	1.070	0.977	1.062	1.022
FYFF	1.101	0.819	0.810	1.048	1.150*	1.127*	FYFF	0.894	0.665	0.816	0.759**	0.833**	0.657*	FYFF	1.312*	0.977	0.965	1.176*	1.290**	1.265**
PUNEW	1.025	1.016	1.016	1.006	0.990	0.991	PUNEW	1.020	1.011	0.977	0.992	0.977	1.011	PUNEW	1.032	1.023	1.023	0.976	0.961	0.962
dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.016	1.120	1.094	1.019	1.011	0.973	IPS10	0.928	1.023	0.999	0.994	0.986	0.949*	IPS10	0.983	1.084	1.058	1.041	1.033	0.994
FYFF	0.969	4.693**	4.637**	0.732*	0.859	0.842	FYFF	0.665**	3.220**	3.182**	0.942	1.107*	1.085*	FYFF	1.043	5.053**	4.993**	0.961	1.129*	1.107
PUNEW	1.094*	1.045	1.045	1.001	1.028	1.028	PUNEW	1.158	1.106	1.106	0.977	1.003	1.004	PUNEW	1.100*	1.051	1.051	0.972	0.998	0.999
dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	0.862	0.976	0.953	0.855*	0.877*	0.844*	IPS10	0.776	0.878	0.858	0.732**	0.751**	0.723*	IPS10	0.818**	0.925	0.903	0.925*	0.949	0.913*
FYFF	1.465***	3.074***	3.037***	0.801**	0.942	0.923	FYFF	1.128	2.366**	2.338**	0.711***	0.836	0.819	FYFF	1.161	2.437**	2.407**	0.819**	0.963	0.944
PUNEW	1.085	0.982	0.982	1.014	0.974	0.974	PUNEW	1.143*	1.034	1.034	0.996	0.956	0.957	PUNEW	1.118**	1.012	1.012	0.985	0.946	0.947
dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.023	0.650**	0.635**	0.984	0.727	0.699	IPS10	1.022	0.650**	0.635**	1.247	0.921	0.886	IPS10						
FYFF	1.586**	3.664**	3.620**	0.989	0.659*	0.646	FYFF	0.952	2.199**	2.173**	1.529	1.020	1.000	FYFF						
PUNEW	0.967	1.004	1.005	1.014	0.953	0.953	PUNEW	1.144**	1.189	1.189	0.994	0.934*	0.934*	PUNEW						
dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.129	0.579**	0.566***	0.957	0.745	0.717	IPS10	1.555***	0.797	0.779	1.032	0.804	0.774	IPS10	1.227	0.629**	0.615**	1.178	0.918	0.883
FYFF	1.189	1.552	1.534	0.964	0.803	0.787	FYFF	0.813	1.062	1.049	1.186*	0.987	0.968	FYFF	1.282*	1.674	1.654	2.033*	1.692**	1.659**
PUNEW	1.198**	0.972	0.972	1.013	0.985	0.985	PUNEW	1.062	0.862	0.862	0.984	0.957	0.957	PUNEW	1.205*	0.978	0.978	1.044	1.015*	1.015*

Hor:1 is the first-step-ahead forecast and Hor:4 is the fourth-step-ahead forecast. We assess forecast accuracy in terms of the sample RelMSFE. We report RelMSFE measures for the three key macroeconomic variables. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and Kronecker indices were specified according to the HK, MT, and OZ algorithms. The symbols *, **, and *** denote rejection, at the 10%, 5%, and 1% levels, of the null hypothesis of equal predictive accuracy according to the Diebold and Mariano (1995) test.

Table 2.13: Forecast: Large Systems

Hannan-Kavalieris (HM) Algorithm K = 20, T = 470, 50 out of sample forecast							MT Algorithm K = 20, T = 470, 50 out of sample forecast							OZ Algorithm K = 20, T = 470, 50 out of sample forecast						
	Hor:1			Hor:4				Hor:1			Hor:4				Hor:1			Hor:4		
dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	2.080**	1.688	1.648	0.966	0.750	0.721	IPS10	1.884*	1.529	1.493	0.966	0.749	0.721	IPS10	1.202	0.975	0.953	2.422*	1.879**	1.809**
FYFF	0.792	3.859***	3.813***	0.883	0.689	0.675	FYFF	0.778	3.791***	3.745***	0.929	0.725	0.711	FYFF	1.004	4.892**	4.833**	1.526	1.190**	1.167**
PUNEW	1.236*	1.169	1.169	1.000	0.955	0.955	PUNEW	1.223*	1.156	1.156	1.006	0.961	0.961	PUNEW	1.135*	1.073	1.074	1.026	0.980	0.980
dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.075	1.509**	1.473*	1.062	1.097	1.056	IPS10	0.956	1.342	1.310	0.665**	0.687*	0.662*	IPS10	0.886	1.244	1.214	0.934	0.965	0.929
FYFF	1.087	4.316**	4.265**	1.064	1.210	1.186	FYFF	0.415*	1.649	1.630	0.538*	0.612	0.600	FYFF	1.194	4.742**	4.685**	0.969	1.102**	1.080**
PUNEW	1.066	1.444	1.444	1.083	1.091	1.092	PUNEW	1.455	1.971	1.971	0.950	0.957	0.958	PUNEW	1.022	1.384	1.384	0.994	1.002	1.003
dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.326	1.205	1.176	1.003	0.892	0.859	IPS10	1.280	1.163	1.136	1.903	1.693*	1.630*	IPS10	0.993	0.903	0.881	0.968	0.861	0.829
FYFF	0.922	3.706***	3.662***	1.225	1.201	1.177	FYFF	1.171	4.711*	4.654*	1.108	1.087	1.065	FYFF	1.087	4.369**	4.317**	1.100	1.079**	1.058**
PUNEW	1.125	1.316	1.316	1.023	0.973	0.974	PUNEW	1.163*	1.361	1.361	1.070	1.018	1.018	PUNEW	1.076	1.259	1.259	1.111	1.057	1.058
dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	0.908	0.921	0.899	0.951	0.862	0.830	IPS10	1.059	1.073	1.048	0.966	0.876	0.843	IPS10	1.024	1.039	1.014	1.054	0.957	0.921
FYFF	1.334*	1.898	1.875	0.650	0.583	0.571	FYFF	1.444**	2.054	2.030	0.939	0.842	0.825	FYFF	1.061	1.509	1.491	1.384	1.241**	1.216**
PUNEW	1.055	0.927	0.927	1.085	1.054	1.054	PUNEW	1.070	0.941	0.941	1.106	1.074	1.074	PUNEW	1.042*	0.916	0.916	1.035	1.005	1.006
dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	2.168**	3.135***	3.061***	0.999	1.155	1.112	IPS10	1.890*	2.732**	2.668**	0.822*	0.951	0.916	IPS10	0.899	1.300	1.269	0.787*	0.911	0.877
FYFF	1.472*	4.170***	4.120***	1.057	1.218	1.194	FYFF	1.497*	4.241***	4.190***	0.808*	0.931	0.913	FYFF	1.072	3.036**	2.999**	0.993	1.145**	1.122**
PUNEW	1.514	2.117**	2.117**	1.039	0.980	0.980	PUNEW	1.895*	2.648**	2.649**	1.060	1.000	1.000	PUNEW	1.016	1.421	1.421	1.085	1.023	1.024

Hor:1 is the first-step-ahead forecast and Hor:4 is the fourth-step-ahead forecast. We assess forecast accuracy in terms of the sample RelMSFE. We report RelMSFE measures for the three key macroeconomic variables. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm and Kronecker indices were specified according to the HK, MT, and OZ algorithms. The symbols *, **, *** denote rejection, at the 10%, 5%, and 1% levels, of the null hypothesis of equal predictive accuracy according to the Diebold and Mariano (1995) test.

Table 2.14: Forecast: Large Systems

Hannan-Kavalieris (HM) Algorithm K = 40, T = 470, 50 out of sample forecast							MT Algorithm K = 40, T = 470, 50 out of sample forecast							OZ Algorithm K = 40, T = 470, 50 out of sample forecast						
	Hor:1			Hor:4				Hor:1			Hor:4				Hor:1			Hor:4		
dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 1	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	5.8095**	7.5537**	7.3534**	2.040	2.050	1.961	IPS10	0.999	1.299	1.264	1.033	1.038	0.993	IPS10	1.088	1.415	1.382	0.917	0.922	0.887
FYFF	1.389	4.9776***	4.9129***	1.248	1.253	1.228	FYFF	1.148	4.1132*	4.0598*	0.992	0.996	0.976	FYFF	1.140	4.085	4.036	1.070	1.075	1.054
PUNEW	0.669	1.188	1.188	1.108	1.062	1.062	PUNEW	0.567	1.006	1.006	1.050	1.006	1.006	PUNEW	1.008	1.789	1.789	1.022	0.980	0.980
dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 2	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.449	1.573	1.532	1.163	1.273	1.218	IPS10	0.996	1.081	1.052	0.927	1.014	0.970	IPS10	1.073	1.165	1.138	0.862	0.943	0.908
FYFF	4.0506**	15.6133***	15.4106***	1.882	2.112	2.070	FYFF	1.227	4.731	4.669	0.908	1.019	0.999	FYFF	1.146	4.417	4.364	0.790	0.886	0.869
PUNEW	1.019	1.372	1.372	1.232	1.116	1.116	PUNEW	0.758	1.021	1.021	1.113	1.008	1.008	PUNEW	0.983	1.324	1.324	1.097	0.994	0.994
dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 3	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	0.952	1.241	1.212	3.416	3.159	3.040	IPS10	0.989	1.290	1.256	1.097	1.014	0.970	IPS10	1.122	1.463	1.429	0.964	0.891	0.858
FYFF	0.872	5.904*	5.834*	1.261	1.032	1.012	FYFF	0.698	4.7259**	4.6645**	1.245	1.019	0.999	FYFF	0.909	6.153*	6.080*	1.266	1.036	1.016
PUNEW	0.986	1.279	1.279	1.628	1.504	1.505	PUNEW	0.788	1.022	1.022	1.091	1.008	1.008	PUNEW	1.049	1.361	1.361	1.122	1.037	1.037
dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 4	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	0.896	0.931	0.909	0.903	0.883	0.849	IPS10	1.011	1.050	1.022	1.384	1.352	1.293	IPS10	1.199	1.245	1.216	1.578	1.542	1.484
FYFF	1.051	4.917	4.858	1.358	1.218	1.194	FYFF	0.959	4.484	4.426	0.870	0.780	0.764	FYFF	0.989	4.626	4.571	1.298	1.164	1.141
PUNEW	1.299	1.344	1.344	1.027	0.952	0.952	PUNEW	0.959	0.992	0.992	1.081	1.002	1.002	PUNEW	0.978	1.012	1.012	1.123	1.041	1.041
dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡	dataset 5	VAR	AR(1)	AR(1)‡	VAR	AR(1)	AR(1)‡
IPS10	1.110	1.373	1.341	0.922	0.910	0.876	IPS10	1.007	1.247	1.214	1.053	1.040	0.995	IPS10	1.189	1.472	1.438	0.897	0.886	0.853
FYFF	1.020	3.633	3.590	1.446	1.356	1.329	FYFF	1.220	4.3455*	4.289*	1.058	0.992	0.972	FYFF	1.071	3.814	3.769	1.052	0.986	0.967
PUNEW	1.183	2.101**	2.102**	0.990	0.930	0.930	PUNEW	0.569	1.010	1.010	1.071	1.007	1.007	PUNEW	1.000	1.775	1.776	1.058	0.995	0.995

Hor:1 is the first-step-ahead forecast and Hor:4 is the fourth-step-ahead forecast. We assess forecast accuracy in terms of the sample RelMSFE. We report RelMSFE measures for the three key macroeconomic variables. RelMSFE measures less than one imply that VARMA models outperform the assigned competitor. We estimate VARMA models with the IOLS algorithm, and Kronecker indices were specified according to the HK, MT, and OZ algorithms. The symbols *, **, and *** denote rejection, at the 10%, 5%, and 1% levels, of the null hypothesis of equal predictive accuracy according to the Diebold and Mariano (1995) test.

Chapter 3

The Nonlinear Iterative Least Squares (NL-ILS) Estimator: An Application to Volatility Models

3.1 Introduction

Measuring volatility and identifying its sources is of major importance in finance and economics. Investors are concerned about asset return volatility because it plays crucial role on asset pricing, risk management and portfolio allocation. As a result, the task of modeling the conditional variance has been a central topic in econometrics following the seminal papers of Engle (1982) and Bollerslev (1986). Since then, different specifications and frameworks, such as GARCH-type models, stochastic volatility, realized volatility and combinations of these approaches have been adopted,

trying to capture the very specific stylized facts observed in financial returns. A natural extension that emerges from modeling the conditional variance is the relation between risk and return. The intertemporal capital asset pricing model (ICAPM) of Merton (1973) establishes a positive relation between the conditional excess returns and the conditional variance, implying that investors should be remunerated for bearing extra risk. Engle, Lilien, and Robins (1987) provide the first econometric specification that relates the conditional second moment to the first moment, allowing to test the ICAPM model. Following them, several attempts have been undertaken to estimate the risk premium parameter, however empirical evidences on the sign and significance of this parameter are blurred. Two potential causes are: firstly, as Bollerslev, Chou, and Kroner (1992) point out, quasi-maximum likelihood (QMLE) estimates of the risk premium parameter using the GARCH-in-mean framework may be inconsistent if the conditional variance is misspecified. Secondly, as Drost and Nijman (1993) discuss, sampling frequency impacts the validity of the assumptions governing the QMLE estimator and as a consequence of time aggregation, estimates of the risk premium parameter may be inconsistent[1].

In this chapter, we address the two above-mentioned issues, by proposing a novel full parametric iterative estimator, the nonlinear iterative least squares estimator (NL-ILS). The NL-ILS estimator nests the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models[2]. We derive the consistency and asymptotic distribution for the

[1] Linton and Perron (2003), Linton and Sancetta (2009), Conrad and Mammen (2008), Christensen, Dahl, and Iglesias (2012) point a third issue. They find strong evidences that the relation between risk and return is nonlinear, indicating that the mixed results obtained with the full parametric GARCH-in-mean models could be the results of misspecification of the mean equation.

[2] Under certain assumptions briefly discussed in Section 3.2.1, the NL-ILS is also valid

GARCH(1,1) model under mild assumptions. The asymptotic results for the NL-ILS estimator do not depend on the correct specification of the stochastic term distribution, allowing, therefore, the NL-ILS estimator to compete against the QMLE estimator. Moreover, we extend the consistency result to the weak-GARCH(1,1) case, which as far as our knowledge goes, is only covered by the estimator proposed by Francq and Zakoian (2000). Furthermore, we show through Monte Carlo exercises that the NL-ILS estimator is more robust to misspecification of the conditional variance than the QMLE estimator when considering the GARCH(1,1)-in-mean case. This result is particularly important when investigating the existence of the risk-return tradeoff, since the true data generation process (DGP) of the conditional variance is unknown in practise. We find evidences that bias on the QMLE estimates of the risk premium parameter leads to false significant risk premium estimates in a full parametric GARCH(1,1)-in-mean model.

The literature on GARCH-type models is extremely extensive, with a wide range of specifications aiming to capture different stylized facts (see Francq and Zakoian (2010) and Bollerslev (2008)). In this chapter, we will focus on the following models: GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean originally proposed by Bollerslev (1986), Drost and Nijman (1993), Engle, Lilien, and Robins (1987) and Hansen, Huang, and Shek (2012), respectively. Another important branch of the GARCH literature examines the asymptotic properties of the QMLE estimator. Research on this topic has mainly focused on relaxing moment assumptions as a way to accommodate heavy-tailed

for ARMA(1,1) and weak-ARMA(1,1) models.

marginal distributions (see Francq and Zakoian (2008) for a survey on this topic). We address this issue by establishing the asymptotic theory for the GARCH(1,1) model under assumptions that are compatible with the QMLE estimator. Apart from Christensen, Dahl, and Iglesias (2012) of which work nests the full parametric GARCH(1,1)-in-mean and which is based on the profile log-likelihood approach, there has not been so far a proper QMLE asymptotic theory covering this model. This chapter discusses the extension of the NL-ILS asymptotic results for the GARCH(1,1)-in-mean and the RealGARCH(1,1)-in-mean models.

Recently, the abundant availability of high frequency data has triggered a new class of volatility models: the realized volatility (see Mcaleer and Medeiros (2008) for an extensive survey on the different estimators available in the literature). Jointly with that, models that combine GARCH-type structure with realized measures, such as GARCH-X in Engle (2002), HEAVY in Shepard and Sheppard (2010) and RealGARCH in Hansen, Huang, and Shek (2012), have also become popular. These "turbo"[3] models have the nice property of adjusting much faster to shocks in volatility, providing a better forecasting performance than GARCH-type models. By extending the NL-ILS algorithm to the RealGARCH(1,1)-in-mean model, we able to assess whether, by augmenting the volatility equation with realized variance measures, the risk premium parameter estimate improves. Moreover, the theoretical framework we use to establish the asymptotic theory for the GARCH(1,1) model can also be extended to accommodate exogenous regressors in the variance equation as in the RealGARCH(1,1)-in-mean model. Another important advantage of the NL-ILS framework

[3]This is an expression used by Shepard and Sheppard (2010), and it illustrates, in a very good way, the enhanced properties of this class of augmented models.

emerges from its robustness to disturbances that possess some nonlinear dependence. From the RealGARCH framework, the measurement equation relates the conditional variance to the realized variance. Hansen, Huang, and Shek (2012) assume the stochastic term in the measurement equation is an independent and identically distributed (*iid*) process evolving on daily basis. We argue the conditional and realized variance evolve at different frequencies. The former one evolves on a daily basis, whereas the latter evolves intradaily. Following that, modeling the stochastic term in the measurement equation as an *iid* process might turn out to be a far too strong assumption. Hence, it makes necessary the adoption of estimators that can cope with disturbances possessing dependence on higher moments, such as linear projections, as discussed in Drost and Nijman (1993) and Drost and Werker (1996).

In the empirical section we investigate the existence of the risk premium in the spirit of the ICAPM model proposed by Merton (1973). To do so, we adopt the GARCH(1,1)-in-mean specification. The main question is whether the risk premium parameter is significant and presents the correct sign by using an estimator which is robust to misspecification of the conditional variance and also to dependence on the errors. We assess this question in two dimensions: temporal frequency and market *proxy*. To evaluate the former one we estimate the model on daily, weekly and monthly basis. To appraise the latter dimension we adopt three market indices: CRSP, S&P500 and S&P100. The choice of comparing different indices emerges from the different compositions they have. The CRSP data set is known to be the best *proxy* for the market. When we implement the NL-ILS, the risk premium is significant only in the CRSP data

set. A different picture arises when we use the QMLE estimator: the risk premium is significant in all frequencies and indices. This result holds across all three frequencies. Following the consistency issue of the QMLE estimator, we perform robustness checks using RealGARCH-in-mean (with both NL-ILS and QMLE estimators), EGARCH-in-mean, GJR-GARCH-in-mean and APARCH-in-mean models. These exercises deliver results for the risk premium estimates which are in line with the ones found when using the robust NL-ILS estimator. We argue that the NL-ILS estimator is able to capture the "true" risk premium, since its results reflect the wider composition of the CRPS index, resembling the market more accurately, when compared to S&P500 and S&P100 indices.

This chapter is organized as follows. Section 3.2 introduces the NL-ILS estimator and establishes the asymptotic theory for the GARCH(1,1) case. We start the discussion with a generic model nesting the GARCH, GARCH-in-mean and RealGARCH models. We then illustrate the specific cases of GARCH(1,1), weak-GARCH(1,1) and GARCH(1,1)in-mean. Section 3.3 presents the NL-ILS algorithm for the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean. Section 3.4 displays an extensive Monte Carlo study, assessing the finite sample performance of the NL-ILS compared to the QMLE benchmark with respect to consistency, efficiency and forecast accuracy. This section also discusses the robustness of the NL-ILS estimator when the conditional variance is misspecified. In Section 3.5, we assess the risk-return tradeoff considering different indices at three sampling frequencies. Section 3.6 concludes. The Appendix contains all proofs.

3.2 Asymptotic theory: main results

This section provides theoretical results regarding the consistency and asymptotic distribution of the NL-ILS estimator. We start with a generic model nesting three of the models we discuss through out this chapter: GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. Firstly, we derive the consistency and asymptotic distribution for this generic model under high level assumptions. Secondly, we relax some of these assumptions focusing on these models and analyzing them in greater depth, providing discussions on the asymptotic results. The theory developed in this section is based on the work of Dominitz and Sherman (2005). Following their work, the crucial point on showing consistency and asymptotic distribution for this class of iterative estimators lies on proving that the population mapping is an Asymptotic Contraction Mapping (ACM)[4]. If the population mapping is an ACM, then it has a fixed point. This allows the use of the fixed point theorem to derive the consistency of the iterative estimator. The asymptotic theory is derived using the population mapping evaluated at the true vector of parameters, allowing the use of asymptotic results obtained from the standard nonlinear least squares (NL-LS) framework.

Assume a stationary stochastic process $\{y_t\}_{t=1}^{T}$ with finite fourth mo-

[4]Using the definition in Dominitz and Sherman (2005), a collection $\{K_T^{\omega}(.) : T \geq 1, \omega \in \Omega\}$ is an ACM on $(\mathbb{B}, d)$ if $d\left(K_T^{\omega}(x), K_T^{\omega}(y)\right) \leq cd(x,y)$ as $T \longrightarrow \infty$, where $c \in [0,1)$, $(\mathbb{B}, d)$ is a metric space with $x, y \in \mathbb{B}$, $(\Omega, \mathcal{A}, \mathcal{P})$ denoting a probability space and $K_T^{\omega}(.)$ is a function defined on $\mathbb{B}$.

ment.

$$y_t = f\left(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1\right) + \epsilon_t, \qquad l \geq 1, \quad m \geq 0 \tag{3.1}$$

$$\epsilon_t = \sigma_t \eta_t \tag{3.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i} + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{r} \gamma_i v_{t-i} \tag{3.3}$$

$$Z_t = \Psi_0 + U_t + \sum_{i=1}^{\infty} \Psi_i U_{t-i} \qquad \Psi_i = \varrho_i\left(\theta_2\right), \quad i = 0, 1, ..., \infty \tag{3.4}$$

where $f\left(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1\right)$ is a twice continuously differentiable function; $\varrho_i\left(\theta_2\right)$ is a continuous function for all i's; X_{t-m} is a matrix of exogenous regressors; Y_{t-l} is a vector containing lags of the dependent variable; σ_t^2 is a latent variable (conditional variance); $Z_t = (\epsilon_t^2, v_t)'$; U_t is a vector of martingale difference sequence (m.d.s.) processes, such that $\mathbb{E}\left(U_t\right) = 0$ and $\text{Var}(U_t) = \Sigma_U$ with Σ_u being a diagonal matrix; θ_1 is a vector of free parameters in (3.1), θ_2 is a vector of free parameters in (3.3) and $\theta = (\theta_1, \theta_2)'$. Denote $\mathbb{B}$ as the space where θ is defined. Equation (3.1) is generic enough to accommodate models that are nonlinear in the parameters, also nesting linear regressions. As in Dominitz and Sherman (2005), we define two mappings: population and sample mappings. Both mappings map from $\mathbb{B}$ to itself, and on each iteration, they are computed through the minimization of the average of squared residuals. For notation purposes, we denote the sample objective function as $Q_T\left(y_t, v_t; \theta\right)$ and its population counterpart as $\mathbb{E}\left(Q_T\left(y_t, v_t; \theta\right)\right)$. These functions are nonlinear in the parameters, yielding NL-LS estimates on all iterations. Therefore, the NL-ILS estimator consists on computing NL-LS estimates using estimates of the latent variables as regressors, updating, at each iteration, the latent variable using the NL-

LS parameter estimates. This procedure is repeated until the parameters converge.

Definition 2 *Mapping:*

Define the population mapping as $N(\theta_j)$ *and its sample counterpart as* $\widehat{N}_T(\widehat{\theta}_j)$*, such that at any* j *iteration,* $N(\theta_j)$ *maps from* θ_j *to* θ_{j+1} *and* $\widehat{N}_T(\widehat{\theta}_j)$ *maps from* $\widehat{\theta}_j$ *to* $\widehat{\theta}_{j+1}$*.*

$$\theta_{j+1} = N(\theta_j) = \min_{\theta_{j+1}} \ \mathbb{E}\left\{ \frac{1}{T}\sum_{t=1}^{T}\left[Z_t - \Psi_{j+1,0} - \sum_{i=0}^{\infty} \Psi_{j+1,i} U_{j,t-1-i} \right]^2 \right\} \quad (3.5)$$

$$\widehat{\theta}_{j+1} = \widehat{N}_T(\widehat{\theta}_j) = \min_{\widehat{\theta}_{j+1}} \ \frac{1}{T}\sum_{t=1}^{T}\left[\widehat{Z}_t - \widehat{\Psi}_{j+1,0} - \sum_{i=0}^{\bar{q}} \widehat{\Psi}_{j+1,i} \widehat{U}_{j,t-1-i} \right]^2 \quad (3.6)$$

Note that $\Psi_{j+1,i}$ and $\widehat{\Psi}_{j+1,i}$ depend on $\theta_{2,j+1}$ and $\widehat{\theta}_{2,j+1}$, respectively. The subscript j denotes the iteration which parameters are computed. The number of iterations j is set to be a function of T, such that as $T \longrightarrow \infty$, $j \longrightarrow \infty$ at some rate satisfying $\frac{\ln(T)}{j} = o(1)$ and $\bar{q}$ is a truncation parameter, such that $\bar{q} \longrightarrow \infty$ at a logarithmic rate of T. From the population mapping definition, $\theta = N(\theta)$ holds as an identification condition. This implies that, if $N(\theta_j)$ is an ACM, then θ is the fixed point of the population mapping and the following bound holds for any j:

$$|\theta_{j+1} - \theta_j| = |N(\theta_j) - N(\theta_{j-1})| \leq \kappa \, |\theta_j - \theta_{j-1}| \quad (3.7)$$

where $\kappa = [0,\ 1\ )$ is the contraction parameter. By using the Newton-Raphson (NR) procedure, the two mappings in Definition 1 have the fol-

lowing linear representation:

$$\theta_{j+1} = N(\theta_j) = \theta_j - [H(\theta_j)]^{-1} G(\theta_j) \tag{3.8}$$

$$\widehat{\theta}_{j+1} = \widehat{N}_T\left(\widehat{\theta}_j\right) = \widehat{\theta}_j - \left[\widehat{H}_T\left(\widehat{\theta}_j\right)\right]^{-1} \widehat{G}_T\left(\widehat{\theta}_j\right) \tag{3.9}$$

where $\widehat{G}_T\left(\widehat{\theta}_j\right)$ and $\widehat{H}_T\left(\widehat{\theta}_j\right)$ are the sample gradient and Hessian computed from $Q_T\left(\widehat{\theta}_j\right)$, whereas $H(\theta_j)$ and $G(\theta_j)$ are their population counterparts. To use the theory developed by Dominitz and Sherman (2005), we introduce assumptions which are related to the identification of classical nonlinear regression models (see Amemiya (1985) pg. 129 for more details), and assumptions governing the behavior of both population and sample mappings.

Assumptions A

1. $$\mathbb{E}\left\{\left[Z_t - \widetilde{Z}_t + \left(\varrho_0(\theta_2) + \sum_{i=1}^{\infty} \varrho_i(\theta_2)\right) - \left(\varrho_0\left(\widetilde{\theta}_2\right) + \sum_{i=1}^{\infty} \varrho_i\left(\widetilde{\theta}_2\right)\right)\right]^2\right\} \neq 0$$

   for $\forall\ \widetilde{\theta} \neq \theta$ and $\widetilde{Z}_t = (\widetilde{\epsilon}_t^2, v_t)'$ and $\widetilde{\epsilon}_t^2 = \left(y_t - f\left(Y_{-l}, X_{t-m}, \widetilde{\sigma}_t, \widetilde{\theta}_1\right)\right)^2$.

2. $\mathrm{Cov}(f(Y_{-l}, X_{t-m}, \sigma_t, \theta_1), \epsilon_t) = 0$.

3. The disturbances η_t have a non-degenerate distribution such that $\eta_t \sim iid(0,1)$ and $\mathbb{E}(\eta_t^4) < \infty$.

4. $N(\theta_j)$ is an ACM in spirit of the definition of Dominitz and Sherman (2005) for all $\theta_j \in \mathbb{B}$.

5. $\sup_{\xi,\varsigma \in \mathbb{B}} \left|N(\xi) - \widehat{N}_T(\varsigma)\right| = o_p(1)$ for all $\xi, \varsigma \in \mathbb{B}$. $\theta_j \in \mathbb{B}$.

Assumption A1 implies the population mapping is identified, allowing the use of the NL-LS estimator to recover estimates of θ. Note that Assumption A2 is weaker than the usual assumption presented in linear regressions with stochastic regressors. In these cases, the regressors at time t are assumed to be independent of ϵ_s for all t and s as discussed in Hamilton (1994) chapter 8. Our setup, however, relies on relaxing this assumption in spirit of the AR(p) model (case 4 in chapter 8 of Hamilton (1994)). Assumption A4 states that population mapping in Definition 1 is an ACM. Assumptions A5 establishes uniform convergence between the sample and population mappings. Under Assumptions A1, A2, A3, A4 and A5, Theorems 2 and 4 in Dominitz and Sherman (2005) hold, yielding the consistency and the asymptotic distribution of the NL-ILS algorithm for the generic model defined in (3.1), (3.3) and (3.4).

3.2.1 GARCH(1,1)

From the seminal papers of Engle (1982) and Bollerslev (1986), a process $\{y_t\}_{t=1}^{T}$ is said to be a strong GARCH(1,1), (GARCH(1,1) hereafter), if the following structure holds:

$$y_t = \epsilon_t = \eta_t \sigma_t \tag{3.10}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3.11}$$

where η_t is an $iid \sim (0, 1)$ and σ_t^2 is the latent conditional variance. Sufficient conditions on the parameters of (3.11) that guarantee the process in (3.10) is second-order stationary are: $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$ (see Francq and Zakoian (2010) for an extensive study on stationarity so-

lutions to GARCH(p,q) models). The conditional variance equation of the GARCH(1,1) model allows an ARMA(1,1) representation on the form of:

$$\epsilon_t^2 = \omega + a\epsilon_{t-1}^2 + u_t + bu_{t-1} \tag{3.12}$$

where $a = (\alpha + \beta)$ and $b = -\beta$ are the autoregressive and moving average parameters, respectively. Denote ϕ as $\phi = (\omega, a, b)'$. The disturbances $u_t = \epsilon_t^2 - \sigma_t^2$ are m.d.s., such that $\mathbb{E}(u_t) = 0$ and $\text{Var}(u_t) = \sigma_u^2$. If the GARCH(1,1) in (3.10) and (3.11) is covariance stationary, then the ARMA(1,1) in (3.12) can be expressed as MA(∞) as:

$$\epsilon_t^2 = \psi_0 + \sum_{i=1}^{\infty} \psi_i u_{t-i} + u_t \tag{3.13}$$

where $\psi_0 = \frac{\omega}{1-a}$, $\psi_i = a^i(a+b)$.

We establish the consistency and asymptotic distribution for the GARCH(1,1) model, by relaxing some of the high level assumptions we imposed to the generic model. Note that the generic model nests the GARCH(1,1) model, by setting $f(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1) = 0$, lags orders p and q to 1 and $\gamma_i = 0$ for all $i = 1, 2, ..., r$. These imply that the VMA in (3.4) reduces to the MA(∞) depicted in (3.13). Using Definition 1, and setting the sample objective function as $Q_T\left(y_t; \widehat{\phi}_{j+1}\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\epsilon_t^2 - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2$, the population and the sample mappings are defined as:

Definition 3 *GARCH(1,1) Mapping:*

Define the population mapping for the GARCH(1,1) model as $N(\theta_j)$ and its sample counterpart as $\widehat{N}_T(\widehat{\theta}_j)$, such that at any j iteration, $N(\theta_j)$ maps

from θ_j *to* θ_{j+1} *and* $\widehat{N}_T(\widehat{\theta}_j)$ *maps from* $\widehat{\theta}_j$ *to* $\widehat{\theta}_{j+1}$.

$$\phi_{j+1} = N(\phi_j) = \min_{\phi_{j+1}} \mathbb{E}\left\{ \frac{1}{T}\sum_{t=1}^{T}\left[\epsilon_t^2 - \psi_{j+1,0} - \sum_{i=0}^{\infty}\psi_{j+1,i}u_{j,t-1-i}\right]^2 \right\} \tag{3.14}$$

$$\widehat{\phi}_{j+1} = \widehat{N}_T\left(\widehat{\phi}_j\right) = \min_{\widehat{\phi}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\epsilon_t^2 - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2 \tag{3.15}$$

Remark Definition 3: the MA(∞) representation satisfies assumption A1 in the generic setup (see Lemma 8 in the appendix). We relax assumption A4 and A5 in order to establish the consistency and asymptotic distribution of the NL-ILS estimator. To this purpose, we state the following assumptions:

Assumptions B

1. The GARCH(1,1) model stated in (3.10) and (3.11) is second-order stationary and yields $\sigma_t^2 > 0$ for all t. These imply that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. Also, assume that $\phi \in \mathbb{B}$ and $\mathbb{B}$ is compact.

2. The disturbances η_t have a non-degenerate distribution such that $\eta_t \sim iid(0,1)$ and $\mathbb{E}(\eta_t^4) < \infty$.

3. Define the gradient of $N(\phi_j)$ as $V(\phi_j) = \bigtriangledown_{\phi_j} N(\theta_j)$ and its sample counterpart as $\widehat{V}_T(\widehat{\phi}_j) = \bigtriangledown_{\widehat{\phi}_j} \widehat{N}_T\left(\widehat{\phi}_j\right)$. Then, the Euclidean norm of $\widehat{V}_T(\widehat{\phi}_j)$ is bounded in probability, such that $\left\|\widehat{V}_T(\widehat{\phi}_j)\right\| = O_p(1)$ for all $\widehat{\phi}_j \in \mathbb{B}$.

Assumption B2 is important in two different aspects: firstly, it implies that u_t in (3.24) is a m.d.s.[5]; secondly, the finite fourth moment is required as

[5]The *iid* assumption on η_t can be relaxed as in the case of the weak-GARCH(1,1) model. This will only affect the asymptotic distribution of the NL-ILS estimator.

a condition to obtain a finite σ_u^2. Finally, Assumption B3 establishes the Lipschitz condition of $\widehat{N}_T\left(\widehat{\phi}_j\right)$. This implies that $\widehat{N}_T\left(\widehat{\phi}_j\right)$ is stochastically equicontinuous. To show consistency of the NL-ILS estimator, the crucial point consists on proving that the population mapping is an ACM. Lemma 7 delivers this result. Lemma 7 is equivalent to Lemma 1 in Chapter 2, but the distinct feature of the mapping leads to different contraction properties associated with the population mapping.

Lemma 7 *Suppose Assumptions B1, B2 and B3 hold. Then, there exist an open ball centered at ϕ with closure $\mathbb{B}$, such that the mapping $N(\phi_j)$ is an ACM on $(\mathbb{B}, d)$, with $\phi_j \in \mathbb{B}$ for all $j > 0$.*

Figure 3.1 displays the maximum eigenvalue associated with different combinations of parameters satisfying Assumption B1. From Lemma 7, the population mapping has a fixed point such that $N(\phi) = \phi$ holds and the following inequality is valid for all iterations:

$$|\phi_{j+1} - \phi_j| = |N(\phi_j) - N(\phi_{j-1})| \leq \kappa \, |\phi_j - \phi_{j-1}| \tag{3.16}$$

Remark Lemma 7: Lemma 7 can also be extended to the ARMA(1,1) case. Note that the eigenvalues of the population mapping gradient evaluated on the true vector of parameters, $V(\phi)$, are given by:

$$\varepsilon = \left[\frac{a+b}{1+b}, \quad \frac{a(a+b)}{1+b}, \quad \frac{a(a+b)}{1+b}\right]' \tag{3.17}$$

Under the ARMA(1,1) model, Assumption B1 is relaxed such that $|a| < 1$ and $|b| < 1$ hold. For all $b > 0$, the eigenvalues associated with (3.17) are smaller than one in absolute value. This result is particulary relevant for

ARMA(1,1) models generated with a positive moving average parameter (close to unity) and a negative autoregressive parameter (potentially close to one in absolute value). Under such combination of parameter values, Chapter 2 shows that the IOLS estimator is not valid, because its population mapping is not an ACM. This implies that the NL-ILS estimator can, alternatively, be used when convergence is not achieved with the IOLS estimator.

To prove the consistency of the NL-ILS estimator, it is necessary to show that the population mapping and the population gradient converge uniformly to their sample counterparts when evaluated at the same points. These are given by Lemmas 9 and 10, respectively. Lemmas 9 is obtained using the fact that $\bar{q} \longrightarrow \infty$ at a logarithmic rate of T and using the weak law of large numbers. Lemma 11 in the appendix shows that the sample mapping is also an ACM, implying that it also has a fixed point, denoted by $\widehat{\phi}$, such that $\widehat{N}_T\left(\widehat{\phi}\right) = \widehat{\phi}$.

With regard to the asymptotic distribution of the NL-ILS estimator, Lemma 12 gives the $\sqrt{T}$ convergence of $\widehat{\phi}_j$ to $\widehat{\phi}$. This is achieved by allowing the number of iterations goes to infinite as $T \longrightarrow \infty$, such that $\frac{\ln(T)}{j} = o(1)$. As in Chapter 2, we use the fact that, when evaluated at the true vector of parameters and $T \longrightarrow \infty$, the lagged disturbances are no longer latent variables. This implies that asymptotic results from the NL-LS estimator can be used in the final bit of the proof.

Define the following quantities: $A = [I - V(\phi)]^{-1}$; $V(\phi)$ is the gradient of the population mapping evaluated on the true vector of parameters ϕ;

$C_0 = \text{plim}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial h_t(\phi)}{\partial \phi}\frac{\partial h_t(\phi)}{\partial \phi'}\right]$; and $h_t(\phi) = \psi_0 - \sum_{i=1}^{\bar{q}}\psi_i u_{t-i}$.

$$A^{-1} = \begin{pmatrix} \frac{1-a}{b+1} & \frac{\left(\left(a^2+a-1\right)b+1\right)\omega}{b+1} & -\frac{(a-1)(a+1)^2\omega}{b+1} \\ 0 & \frac{\left(a^3-2a+1\right)b}{b+1} & -\frac{\left(a^2-1\right)^2}{b+1} \\ 0 & \frac{(ab+1)^2}{b+1} & \frac{-(ab+2)a^2+b+2}{b+1} \end{pmatrix} \tag{3.18}$$

$$C_0 = \begin{pmatrix} \frac{1}{(1-a)^2} & -\frac{\omega}{(1-a)^3} & 0 \\ -\frac{\omega}{(1-a)^3} & \frac{\omega^2}{(1-a)^4} + \sum_{i=0}^{\bar{q}} d_i^2\sigma_u^2 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 \\ 0 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 & \sum_{i=0}^{\bar{q}} a^{2i}\sigma_u^2 \end{pmatrix} \tag{3.19}$$

The consistency and asymptotic distribution of the NL-ILS is therefore given by:

Theorem 2 *Suppose Assumptions B1, B2 and B3 hold. Then*

i. $\left|\widehat{\phi} - \phi\right| = o_p(1)$ *as* $j \longrightarrow \infty$ *with* $T \longrightarrow \infty$.

ii. $\sqrt{T}\left[\widehat{\phi} - \phi\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2 A C_0^{-1} A'\right)$ *as* $T \longrightarrow \infty$ *and* $\frac{\ln(T)}{j} = o(1)$.

The proof of Theorem 2 is given in the Appendix. The asymptotic covariance matrix can be computed replacing the true vector of parameters with the consistent NL-ILS estimates of ϕ. From item (ii) in Theorem 2, it is clear that the asymptotic variance of the NL-ILS is in fact an augmented version of the one obtained from the NL-LS estimator when all the regressors are fully observed. The closed form for the asymptotic variance of the NL-ILS estimator considering the parameters of GARCH(1,1) in its

original form can be easily obtained. Denote $\Sigma = AC_0^{-1}A'$, $\Sigma_{i,j}$ as the i,j element of Σ and $\theta = [\omega, \alpha, \beta]'$. Then, the asymptotic distribution of θ is given by:

$$\sqrt{T}\left[\widehat{\theta} - \theta\right] \xrightarrow{d} \mathcal{N}\left(0,\, \sigma_u^2\, \Upsilon\right) \tag{3.20}$$

$$\text{with} \quad \Upsilon = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} + \Sigma_{1,3} & -\Sigma_{1,3} \\ \Sigma_{1,2} + \Sigma_{1,3} & \Sigma_{2,2} + \Sigma_{3,3} + 2\Sigma_{2,3} & -\Sigma_{3,3} - \Sigma_{2,3} \\ -\Sigma_{1,3} & -\Sigma_{3,3} - \Sigma_{2,3} & \Sigma_{3,3} \end{pmatrix}$$

Note that the crucial point to prove item (i) in Theorem 2 is the ACM condition of the population mapping, which does not require u_t to be a m.d.s.. In fact, by relaxing Assumption B2 and substituting it by some weaker condition, such that u_t is a linear projection, we establish the consistency of the weak-GARCH(1,1) model.

Corollary 1 *Suppose Assumptions B1 and B3 hold. If*

i. ϵ_t *is a fourth-order stationary white noise process, such that* u_t *in (3.12) is a linear innovation with* $u_t \sim (0, \sigma_u^2)$ *and* $Cov\left(u_t, \epsilon_{t-l}^2\right)\ \forall\ l > 0$ *;*

Then, $\left|\widehat{\phi} - \phi\right| = o_p(1)$ *as* $j \longrightarrow \infty$ *with* $T \longrightarrow \infty$.

3.2.2 GARCH(1,1)-in-mean

To explore the relation between risk and return, Engle, Lilien, and Robins (1987) proposed the ARCH-in-mean model. Following them, a process $\{y\}_{t=1}^{T}$ is said to be a GARCH(1,1)-in-mean model if the structure below

holds:

$$y_t = \lambda\sigma_t + \epsilon_t \tag{3.21}$$

$$\epsilon_t = \eta_t\sigma_t \tag{3.22}$$

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.23}$$

where η_t be an $iid \sim (0,1)$ sequence and σ_t^2 is the latent conditional variance. The parameter λ is usually known as the risk premium parameter. The GARCH(1,1)-in-mean is second-order stationary provided that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$ hold. Similarly to the GARCH(1,1) model, (3.23) allows an ARMA(1,1) representation as:

$$\epsilon_t^2 = \omega + a\epsilon_{t-1}^2 + u_t + bu_{t-1} \tag{3.24}$$

where $a = (\alpha + \beta)$ and $b = -\beta$. We denote $\phi = (\omega, a, b)'$ and $\theta = (\lambda, \phi')'$. The generic model in (3.1), (3.3) and (3.4) nests the GARCH(1,1)-in-mean specification by setting $f\left(Y_{t-l}, X_{t-m}, \sigma_t, \theta_1\right) = \lambda\sigma_t^2$ and $\gamma_i = 0$ for all $i = 1, 2, ...r$. Extension of Theorem 2 to the GARCH(1,1)-in-mean model does not carry any significant difference. The main point consists on showing that the gradient associated with the population mapping is an ACM. This chapter provides numerical evidences indicating that the gradient of the population mapping is indeed an ACM.

Definition 4 *GARCH(1,1)-in-mean Mapping:*
Define the population mapping for the GARCH(1,1)-in-mean model as $N(\theta_j)$ *and its sample counterpart as* $\widehat{N}_T(\widehat{\theta}_j)$*, such that at any* j *iteration,* $N(\theta_j)$ *maps from* θ_j *to* θ_{j+1} *and* $\widehat{N}_T(\widehat{\theta}_j)$ *maps from* $\widehat{\theta}_j$ *to* $\widehat{\theta}_{j+1}$.

$$\theta_{j+1} = N(\theta_j) = \min_{\phi_{j+1}} \mathbb{E}\left\{ \frac{1}{T}\sum_{t=1}^{T}\left[[y_t - \lambda_{j+1}\sigma_{j,t}]^2 - \psi_{j+1,0} - \sum_{i=0}^{\infty}\psi_{j+1,i}u_{j,t-1-i}\right]^2 \right\} \tag{3.25}$$

$$\widehat{\theta}_{j+1} = \widehat{N}_T\left(\widehat{\theta}_j\right) = \min_{\widehat{\theta}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\left[y_t - \widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right]^2 - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2 \tag{3.26}$$

It is possible to split the sample mapping in two distinct procedures: the first mapping delivers estimates of λ, whereas the second one delivers the parameters from the ARMA(1,1) representation in (3.24). This result is formalized in Proposition 1.

Proposition 1 *Assume the model stated in (3.21), (3.22) and (3.23). Define the vector of free parameters in (3.24) on the $j+1$ iteration as $\widehat{\phi}_{j+1} = \left(\widehat{\omega}_{j+1}, \widehat{a}_{j+1}, \widehat{b}_{j+1}\right)'$. The sample mapping in (3.26) can be computed in two distinct procedures, such that:*

i. $\widehat{\lambda}_{j+1} = \left[\sum_{t=1}^{T}\widehat{\sigma}_{j,t}^2\right]^{-1}\sum_{t=1}^{T}\widehat{\sigma}_{j,t}y_t$

ii. $\widehat{\phi}_{j+1} = \min_{\widehat{\phi}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\left[y_t - \widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right]^2 - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2$

Remark Proposition 1: it provides the necessary identification conditions for the use of NL-ILS estimator and alleviates the computational burden of computing parameter(s) in the mean equation. The proof of Proposition 1 is obtained from the first order condition computed from the sample mapping in Definition 4. Figure 3.2 displays the maximum eigenvalue associated

with the numerical gradient of the sample mapping computed using results in Proposition 1. As discussed in Lemma 7, if the maximum eigenvalue is smaller than one in absolute value, this guarantees the ACM property. All the eigenvalues in Figure 3.2 are less than one in absolute value, indicating that the sample mapping is an ACM. Furthermore, preliminary calculations show that the contraction property of $N(\theta)$ does not depend on λ, being only governed by the parameters from the ARMA(1,1) representation of (3.23). If the ACM holds, asymptotic theory for the GARCH(1,1)-in-mean model can be extended following the steps in Theorem 2.

3.3 NL-ILS estimation procedure

We first describe the NL-ILS algorithm for the simple GARCH(1,1) model. As a natural extension of this procedure, we show that NL-ILS estimator can be also applied to the weak-GARCH(1,1) models. This variant of the GARCH(1,1) model was originally proposed by Drost and Nijman (1993) and it is robust to temporal aggregation. Secondly, we extend the algorithm for the GARCH(1,1)-in-mean model in the spirit of Engle, Lilien, and Robins (1987). This is a particulary interesting case since, under this specification, the mean equation has now a latent regressor. Finally, we show that the NL-ILS algorithm can also be implemented to estimate parameters from the RealGARCH(1,1)-in-mean model in the spirit of Hansen, Huang, and Shek (2012). This model turns out to be particularly important, because it is parameterized in such a way that there is a measurement equation linking the latent conditional variance to the realized measure. Hence, the RealGARCH model can be seen as an augmented GARCH

model.

3.3.1 GARCH(1,1) and weak GARCH(1,1) models

We consider the GARCH(1,1) model as in (3.10) and (3.11). Using the sample mapping defined in (3.15), the NL-ILS algorithm is computed through the following steps:

Step 1: Given any initial estimate[6] of ϕ, denoted by $\widehat{\phi}_0$ with $\widehat{\phi}_0 \in \mathbb{B}$, compute, recursively, estimates of u_t, denoted by $\widehat{u}_{0,t}$, with:

$$\widehat{u}_{0,t} = \epsilon_t^2 - \widehat{\omega}_0 - \widehat{a}_0 \epsilon_{t-1}^2 - \widehat{b}_0 \widehat{u}_{0,t-1} \tag{3.27}$$

Step 2: Plug $\widehat{u}_{0,t}$ into the sample mapping and minimize the sum of squared residuals with respect to $\widehat{\phi}_1$ to obtain the first estimate of ϕ, denoted by $\widehat{\phi}_1$.

$$\widehat{\phi}_1 = \widehat{N}_T\left(\widehat{\phi}_0\right) = \min_{\widehat{\phi}_1} \frac{1}{T}\sum_{t=1}^{T}\left[\epsilon_t^2 - \widehat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{1,i}\widehat{u}_{0,t-1-i}\right]^2 \tag{3.28}$$

where $\widehat{\psi}_{1,0}$ and $\widehat{\psi}_{1,i}$ denote the parameters from the $MA(\infty)$ representation computed using $\widehat{\phi}_1 = (\widehat{\omega}_1, \widehat{a}_1, \widehat{b}_1)'$ and $\bar{q}$ is the truncation parameter defined exogenously. Note that (3.28) arises from the $MA(\infty)$ representation of the conditional variance, following the fact the the conditional variance allows an ARMA representation and assumption B1 guarantees invertibility of the autoregressive polyno-

[6]The starting value $\widehat{\phi}_0$ can assume any value, provided that $\widehat{\phi}_0 \in \mathbb{B}$, where $\mathbb{B}$ is the set of parameters satisfying the restrictions that guarantee the GARCH(1,1) model in (3.10) is second-order stationary. In both Monte Carlo study and empirical analysis, we obtain ϕ_0 by estimating (3.12) using residuals obtained from an AR(p) model as regressors.

mial. The Monte Carlo simulations showed that the size of $\bar{q}$ does not play a decisive role on both performance and convergence. As a standard rule, we fixed $\bar{q} = 3\sqrt[4]{T}$ [7].

Step 3: Compute recursively a new set of residuals, denoted by $\widehat{u}_{1,t}$, using $\widehat{\phi}_1$ through (3.29):

$$\widehat{u}_{1,t} = \epsilon_t^2 - \widehat{\omega}_1 - \widehat{a}_1\epsilon_{t-1}^2 - \widehat{b}_1\widehat{u}_{1,t-1} \tag{3.29}$$

Repeat steps 2 and 3 j times until $\widehat{\phi}_j$ converges. We assume NL-ILS algorithm converges if $\left\|\widehat{\phi}_j - \widehat{\phi}_{j-1}\right\| < c$, where c is exogenously defined. In both, Monte Carlo simulations and empirical application, we set $c = 10^{-5}$. Therefore, the j^{th} iteration of the NL-ILS algorithm is given by the minimization below:

$$\widehat{\phi}_j = \widehat{N}_T\left(\widehat{\phi}_{j-1}\right) = \min_{\widehat{\phi}_j} \frac{1}{T}\sum_{t=1}^{T}\left[\epsilon_t^2 - \widehat{\psi}_{j,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{j,i}\widehat{u}_{j-1,t-1-i}\right]^2 \tag{3.30}$$

We denote the NL-ILS estimates obtained through the steps above by $\widehat{\phi}$. The key factor that guarantees the NL-ILS algorithm converges is the contraction property yielded by the ACM condition on the population counterpart of (3.30). It is important to point out that the speed of convergence depends on the contraction parameter κ as discussed in Section 3.2. Considering a specification such that $\alpha = 0.025$ and $\beta = 0.95$, the maximum eigenvalue associated with $V(\phi)$ is equal to 0.5. Adopting $c = 10^{-5}$ and provided that $\left|\widehat{\phi}_0 - \phi\right| = 0.1$, convergence in this scenario would occur af-

[7] Note that, under the true vector of parameters, the disturbances are *iid* process, implying, from the Theorem 3.1 (Orthogonal Regression) in Greene (2008) - pg. 23, that estimates of ψ are unbiased for any truncation parameter $\bar{q}$. At any iteration j, we only require the residuals to be uncorrelated and the resulting MA($\bar{q}$) model to be invertible.

ter ten iterations. This is in line with the results obtained in the Monte Carlo study, where convergence for the GARCH(1,1) takes, on average, eight iterations.

The class of GARCH(1,1) model suffers from an important drawback: it is not closed under temporal aggregation. To overcome this issue, Drost and Nijman (1993) introduced the weak-GARCH(p,q) model. From their definition, a weak-GARCH(1,1) model, at some frequency m, is given by:

$$y_{(m)t} = \epsilon_{(m)t} = \eta_{(m)t}\sigma_{(m)t} \tag{3.31}$$

$$\sigma^2_{(m)t} = \omega_{(m)} + \alpha_{(m)}\epsilon^2_{(m)t-1} + \beta_{(m)}\sigma^2_{(m)t-1} \tag{3.32}$$

$$\mathbb{E}\left(\epsilon_{(m)t}\right) = 0 \tag{3.33}$$

$$\mathbb{P}\left[\epsilon^2_{(m)t} \mid \epsilon_{(m)t-1}, \epsilon_{(m)t-1}, ...\right] = \sigma^2_{(m)t} \tag{3.34}$$

where $\mathbb{P}\left[\epsilon^2_{(m)t} \mid \epsilon_{(m)t-m}, \epsilon_{(m)t-2m}, ...\right]$ denotes the best linear predictor of $\epsilon^2_{(m)t}$ in terms of the lagged values of $\epsilon_{(m)t}$. An alternative definition of weak-GARCH (in terms of ARMA representation) is given by Francq and Zakoian (2010). They state that a process $\epsilon_{(m)t}$ is generated by a weak-GARCH if $\epsilon_{(m)t}$ is a white noise and $\epsilon^2_{(m)t}$ admits an ARMA representation, such that $u_{(m)t}$ in the ARMA(1,1) representation is a linear innovation with $\text{Cov}\left(u_{(m)t}, \epsilon^2_{(m)t-l}\right)$ for all $l > 0$. By being closed under temporal aggregation, the weak-GARCH(1,1) model relaxes the assumption on sampling the data at the true data generation process frequency. This is particulary relevant when dealing with financial returns which are discrete representations from continuous processes. Drost and Werker (1996) establish the temporal aggregation, from the continuous time processes to the weak-GARCH models, providing closed solutions for the diffusion parameters as functions

of the parameters of the weak-GARCH(1,1) model.

In statistical terms, the main difference between the weak-GARCH and the strong GARCH approaches arises from the nature of the innovations associated with the ARMA representation of the conditional variance. Under the weak-GARCH specification, these innovations are no longer a m.d.s. process as they are in the case of the strong GARCH. This different feature implies that the disturbances of the ARMA representation of the weak-GARCH model may carry some nonlinear dependence. In terms of economic intuition, the weak-GARCH class of model turns to be much more flexible and generic than its strong counterpart, having as a core benefit the fact that it is closed under temporal aggregation.

Since this chapter focuses on discrete time models, we restrict our analysis to the temporal aggregation provided by Drost and Nijman (1993). They define a bridge from the parameters of the strong GARCH(1,1) to the parameters of the weak-GARCH(1,1) sampled at some lower frequency m as the solution of the following system of equations:

$$\omega_{(m)} = \omega \left[\frac{1 - (\beta + \alpha)^m}{1 - (\beta + \alpha)} \right] \tag{3.35}$$

$$\alpha_{(m)} = (\beta + \alpha)^m - \beta_{(m)} \tag{3.36}$$

$$\frac{\beta_{(m)}}{1 + \beta_{(m)}^2} = \frac{\beta (\beta + \alpha)^{m-1}}{1 + \alpha^2 \left[\frac{1-(\beta+\alpha)^{2m-2}}{1-(\beta+\alpha)} \right] + \beta^2 (\beta + \alpha)^{2m-2}} \tag{3.37}$$

where $\omega_{(m)}$, $\alpha_{(m)}$ and $\beta_{(m)}$ are parameters at some frequency m from the weak-GARCH(1,1) model and $\phi_{(m)} = (\omega_{(m)}, \alpha_{(m)}, \beta_{(m)})'$.

Estimation of weak-GARCH models showed to be more difficult than its strong counterpart. In fact, the QMLE asymptotic theory for the weak-GARCH models remain to be established. Monte Carlo exercises, however,

show that QMLE consistently estimate the parameters in (3.32). Francq and Zakoian (2000) establish the asymptotic theory for a two-stage lest squares (LS) estimator.

Extending the NL-ILS algorithm to the weak-GARCH(1,1) is straight forward. Similar to the GARCH(1,1) model, (3.32) allows an ARMA(1,1) representation, implying that steps 1, 2 and 3 above can be performed to obtain estimates of $\phi_{(m)}$. The NL-ILS estimate of $\phi_{(m)}$ is denoted by $\widehat{\phi}_{(m)}$. Corollary 1 delivers the consistency of the NL-ILS for the weak-GARCH(1,1) model. Convergence of the NL-ILS algorithm in the weak-GARCH(1,1) model is the same as in the GARCH(1,1) case, because they both share the same contraction parameter.

3.3.2 GARCH(1,1)-in-mean

We illustrate the NL-ILS estimator for the GARCH(1,1)-in-mean[8] model. From (3.21), (3.22), (3.23) and (3.24), rewrite the model as:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.38}$$

$$\epsilon_t = y_t - \lambda\sigma_t \tag{3.39}$$

$$u_t = \epsilon_t^2 - \omega - a\epsilon_{t-1}^2 - bu_{t-1} \tag{3.40}$$

[8]Note that under the temporal aggregation discussion in the previous subsection, the GARCH(1,1)-in-mean model depicted in (3.21), (3.22) and (3.23) is classified within the strong class of models. The literature, as far as we are aware, does not provide time aggregation results for the GARCH(1,1)-in-mean models. In this entire chapter, therefore, we will refer to the strong GARCH(1,1)-in-mean as GARCH(1,1)-in mean.

Following the fact that $|a| < 1$, the ARMA representation of the conditional variance can be inverted generating an $MA(\infty)$ as:

$$[y_t - \lambda\sigma_t]^2 = \epsilon_t^2 = \psi_0 + \sum_{i=1}^{\infty} \psi_i u_{t-i} + u_t$$

where $\psi_0 = \frac{\omega}{1-a}$, $\psi_i = a^i(a+b)$. The NL-ILS algorithm is computed through the following steps:

Step 1: Choose an initial estimate of θ, such that $\theta_0 \in \mathbb{B}$, where $\mathbb{B}$ is the set of parameters satisfying the second-order stationarity conditions [9]. Using (3.39) and (3.38), compute recursively estimates of the conditional variance, denoted as $\widehat{\sigma}_{0,t}^2$, and estimates of u_t, denoted by $\widehat{u}_{0,t}$.

Step 2: From Proposition 1, the sample mapping in (3.26) can be split in two distinctive maps, such that $\widehat{\theta}_1$ is given by:

$$\widehat{\lambda}_1 = \left[\widehat{\sigma}_0'\widehat{\sigma}_0\right]^{-1}\widehat{\sigma}_0'Y \tag{3.41}$$

$$\widehat{\phi}_1 = \min_{\widehat{\phi}_1} \frac{1}{T}\sum_{t=1}^{T}\left[\left[y_t - \widehat{\lambda}_1\widehat{\sigma}_{0,t}\right]^2 - \widehat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{1,i}\widehat{u}_{0,t-1-i}\right]^2 \tag{3.42}$$

where $\widehat{\sigma}_0$ and Y are $(T \times 1)$ vectors stacking all observations of $\widehat{\sigma}_{0,t}$ and y_t, respectively. Compute $\widehat{\lambda}_1$ through (3.41). Plug $\widehat{\lambda}_1$ into (3.42), and minimize with respect to $\widehat{\phi}_1$. Equations (3.41) and (3.42) deliver $\widehat{\theta}_1$.

Step 3: Using $\widehat{\theta}_1$, compute recursively $\widehat{\sigma}_{1,t}^2$, $\widehat{\epsilon}_{1,t}$ and $\widehat{u}_{1,t}$ through (3.38),

[9] In practise, we firstly fix $\lambda_0 = \left[\frac{1}{T}\sum_{t=1}^{T} y_t\right]\left[Var(y_t)\right]^{-1}$ and obtain $\widehat{\epsilon}_{0,t}$. As a second step, similarly to the GARCH(1,1) case, we estimate an AR(p) model having $\widehat{\epsilon}_{0,t}^2$ as dependent variable. This allows me to get initial estimates of u_t and hence compute $\widehat{\phi}_0$.

(3.39) and (3.40).

Repeat Steps 2 and 3 j times until $\widehat{\theta}_j$ converges. As in the GARCH(1,1) case, we assume $\widehat{\theta}_j$ converges if $\left\|\widehat{\theta}_j - \widehat{\theta}_{j-1}\right\| < c$ [10]. On the j^{th} iteration, the sample mapping resumes to:

$$\widehat{\lambda}_j = \left[\widehat{\sigma}'_{j-1}\widehat{\sigma}_{j-1}\right]^{-1}\widehat{\sigma}'_{j-1}Y \tag{3.43}$$

$$\widehat{\phi}_j = \min_{\widehat{\phi}_j} \frac{1}{T}\sum_{t=1}^{T}\left[\left[y_t - \widehat{\lambda}_j\widehat{\sigma}_{j-1,t}\right]^2 - \widehat{\psi}_{j,0} - \sum_{i=0}^{\bar{q}}\widehat{\psi}_{j,i}\widehat{u}_{j-1,t-1-i}\right]^2 \tag{3.44}$$

Denote the resulting NL-ILS estimates as $\widehat{\theta}$. As in the GARCH(1,1) case, the key factor governing convergence of the NL-ILS estimator is the ACM property of $N_T(\theta_j)$. It is important to stress that the procedure described above covers models with mean equation that accounts for more complex mean specifications. Among them, (3.39) can contain a constant, observed exogenous regressors or any function from the conditional variance. For all these scenarios, Proposition 1 holds, implying that sample mapping, on any j^{th} iteration, can take the form of equations (3.43) and (3.44). With respect to computational issues, the NL-ILS estimator does not present significant numerical problems, achieving convergence for the great majority of replications. To this point, we found that by imposing constraints on the autoregressive and moving average parameters in (3.40), the rate of success of the NL-ILS algorithm improves. This showed to be a valid strategy when dealing with small samples and conditional variance specifications containing β very close to 1.

[10] As in the GARCH(1,1) case, we fix $c = 10^{-5}$.

3.3.3 RealGARCH(1,1)-in-mean

The availability of high frequency data has triggered a new class of volatility estimators: the realized variance. These realized measures showed to be quite powerful on modeling the unobserved conditional variance. As pointed out by Andersen, Bollerslev, Diebold, and Labys (2003), realized volatility measures are able to respond faster to abrupt changes in the underline volatility than the standard GARCH framework, which may deliver massive improvements on volatility forecast. This important feature carried by the realized measures led to the establishment of models that combine the GARCH-type approach with the realized variance. Among these models, we point out the GARCH-X, HEAVY and RealGARCH proposed by Engle (2002), Shepard and Sheppard (2010) and Hansen, Huang, and Shek (2012), respectively. In this chapter, we focus on extending the NL-ILS estimator to the case of the RealGARCH(1,1)-in-mean model [11], since these models present the nice feature of having a measurement equation relating the realized variance to the latent conditional variance. The measurement equation accommodates the measurement error that arises from the difference between the realized variance and the latent conditional variance, as pointed out by Asai, Mcaleer, and Medeiros (2011). This chapter does not address the different realized variance estimators, (an extensive survey can be found on Mcaleer and Medeiros (2008)), we simply assume that the realized measures are obtained through consistent estimators and therefore are treated as observed variables. Following Hansen, Huang, and Shek (2012),

[11] The extension of the NL-ILS estimator to the RealGARCH(1,1) model is a natural simplification of the RealGARCH(1,1)-in-mean case. Hence, this chapter covers the latter one as a way to illustrate the use of the NL-ILS estimator.

a log-linear RealGARCH(1,1)-in-mean model is given by:

$$y_t = \lambda\sigma_t + \epsilon_t \tag{3.45}$$

$$\epsilon_t = \eta_t\sigma_t \tag{3.46}$$

$$\ln\sigma_t^2 = \omega + \beta\ln\sigma_{t-1}^2 + \gamma\ln\nu_{t-1} \tag{3.47}$$

$$\ln\nu_t = \xi + \varphi\ln\sigma_t^2 + \tau\left(\eta_t\right) + z_t \tag{3.48}$$

where ν_t accounts for the realized variance, $\tau\left(\eta_t\right)$ is a leverage function capturing asymmetries on the response of the realized measure to positive or negative shocks in η_t, η_t and z_t are a iid processes with zero mean and variances equal to 1 and σ_z^2 respectively. As in Hansen, Huang, and Shek (2012)), we define the leverage function as:

$$\tau\left(\eta_t\right) = \tau_1\eta_t + \tau_2\left(\eta_t^2 - 1\right) \tag{3.49}$$

Note that (3.48) is a measurement equation relating σ_t^2 to ν_t. Its importance is twofold: firstly, it allows multi-step-ahead forecast, since the dynamics of ν_t is fully specified; secondly, it helps identifying the parameters in (3.45), (3.47) and (3.48) when NL-ILS estimator is adopted. Denote $\theta = (\lambda, \omega, \beta, \gamma, \xi, \varphi, \tau_1, \tau_2)'$. Estimation of the log-linear RealGARCH(1,1)-in-mean is originally undertaken through QMLE procedure. The QMLE estimates are denoted by $\widehat{\theta}_Q$). Hansen, Huang, and Shek (2012) discuss the asymptotic properties of $\widehat{\theta}_Q$ for the standard log-linear RealGARCH(1,1) case.

Hansen, Huang, and Shek (2012) derives a VARMA(1,1) representation of the two processes: $\ln \epsilon_t^2$ and $\ln \nu_t$:

$$\begin{pmatrix} \ln \nu_t \\ \ln \epsilon_t^2 \end{pmatrix} = \begin{pmatrix} \mu_\nu \\ \mu_\epsilon \end{pmatrix} + \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix} \begin{pmatrix} \ln \nu_{t-1} \\ \ln \epsilon_{t-1}^2 \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix} + \begin{pmatrix} -\beta & 0 \\ \gamma & -\rho \end{pmatrix} \begin{pmatrix} w_{t-1} \\ u_{t-1} \end{pmatrix} \tag{3.50}$$

where $\mu_\epsilon = \omega + \gamma\xi + (1 - \beta - \varphi\gamma)\,\bar{\eta}^2$, $\mu_\nu = \varphi\omega + (1-\beta)\,\xi$, $\rho = \beta + \varphi\gamma$, $\bar{\eta}^2 = \mathbb{E}(\ln \eta_t^2)$, $w_t = \tau(\eta_t) + z_t$ and $u_t = \ln \eta_t^2 - \bar{\eta}^2$. Note that $\ln \nu_t$ in (3.50) does not depend on the $\ln \epsilon_t^2$ nor on u_t, implying that the parameters on the first equation of the VARMA(1,1) representation can be estimated separately from the parameters governing $\ln \nu_t$. Furthermore, necessary condition an ARMA(1,1) model to be second-order stationarity implies $|\rho| < 1$. Using this result, (3.50) can be expressed as a VMA(∞) process:

$$\begin{pmatrix} \ln \nu_t \\ \ln \epsilon_t^2 \end{pmatrix} = \begin{pmatrix} \frac{\mu_\nu}{1-\rho} \\ \frac{\mu_\epsilon}{1-\rho} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix} + \begin{pmatrix} \sum_{i=0}^{\infty} \rho^i (\rho - \beta) L^i & 0 \\ \sum_{i=0}^{\infty} \gamma \rho^{i+1} L^i & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ u_{t-1} \end{pmatrix} \tag{3.51}$$

Note that the generic model in (3.1), (3.3) and (3.4) nests the RealGARCH(1,1)-in-mean by setting the mean equation as $\lambda\sigma_t$; α, β and γ to one; and (3.51) satisfies (3.4), where $Z_t = (\ln \nu_t, \ln \epsilon_t^2)'$ and $U_t = (w_t, u_t)'$. The algorithm we adopt uses the fact that the rows in the $VMA(\infty)$ depicted in (3.51) can be estimated separately, solving the minimization problems considering

the free parameters in (3.51). Hence, we minimize the following expressions on each iterations, using estimates of the latent regressors:

$$\widehat{\phi}_{j+1} = \widehat{N}_T\left(\widehat{\phi}_j\right) = \min_{\widehat{\phi}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\ln \nu_t - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{j+1,i}\widehat{w}_{j,t-1-i}\right]^2$$

$$\widehat{\zeta}_{j+1} = \widehat{M}_T\left(\widehat{\zeta}_j\right) = \min_{\widehat{\zeta}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\ln\left[\left(y_t - \widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right)^2\right] - \frac{\widehat{\mu}_{\epsilon,j+1}}{1-\widehat{\rho}_{j+1}} - \sum_{i=0}^{\bar{q}} \widehat{\gamma}_{j+1}\widehat{\rho}_{j+1}^{i}\widehat{u}_{j,t-1-i}\right]^2$$

where $\widehat{\psi}_{j+1,0} = \widehat{\mu}_{\nu,j+1}/\left(1-\widehat{\rho}_{j+1}\right)$ and $\widehat{\psi}_{j+1,i} = \widehat{\rho}_{j+1}^{i}\left(\widehat{\rho}_{j+1} - \widehat{\beta}_{j+1}\right)$, $\widehat{\phi}_{j+1} = \left(\widehat{\mu}_{\nu,j+1}, \widehat{\rho}_{j+1}, \widehat{\beta}_{j+1}\right)'$ and $\widehat{\zeta}_{j+1} = \left(\widehat{\lambda}_{j+1}, \widehat{\mu}_{\epsilon,j+1}, \widehat{\gamma}_{j+1}\right)'$.

Considering the baseline log-linear RealGARCH(1,1)-in-mean model and the compact representation of the $(\ln \epsilon_t^2, \ln \nu_t)'$ in (3.51), the NL-ILS is computed through the following steps:

Step 1: Choose an initial estimate of θ, such that $\theta_0 \in \mathbb{B}$, where $\mathbb{B}$ is the set of parameters satisfying the second-order stationarity conditions. Compute initial estimates of μ_ν and ρ, denoted as $\widehat{\mu}_{\nu,0}$ and $\widehat{\rho}_0$ respectively. Denote $\widehat{\phi}_0 = \left(\widehat{\mu}_{\nu,0}, \widehat{\rho}_0, \widehat{\beta}_0\right)'$ as the vector of parameters describing the $\ln \nu_t$ process. Compute recursively an initial set of disturbances $\widehat{w}_{0,t}$ using:

$$\widehat{w}_{0,t} = \ln \nu_t - \widehat{\mu}_{\nu,0} - \widehat{\rho}_0 \ln \nu_{t-1} + \widehat{\beta}_0 \widehat{w}_{0,t-1} \tag{3.52}$$

Step 2: Recursively in (3.47), compute $\widehat{\sigma}_{0,t}^2$ assuming $\widehat{\theta}_0$.

Step 3: By truncating, at some lag order $\bar{q}$, the first row in (3.51), write

the first sample mapping similarly to the GARCH(1,1) case as:

$$\widehat{\phi}_1 = \widehat{N}_T\left(\widehat{\phi}_0\right) = \min_{\widehat{\phi}_1} \frac{1}{T}\sum_{t=1}^{T}\left[\ln \nu_t - \widehat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{1,i}\widehat{w}_{0,t-1-i}\right]^2 \tag{3.53}$$

where $\widehat{\psi}_{1,0} = \widehat{\mu}_{\nu,1}/\left(1-\widehat{\rho}_1\right)$ and $\widehat{\psi}_{1,i} = \widehat{\rho}_1^i\left(\widehat{\rho}_1 - \widehat{\beta}_1\right)$. Minimize (3.53) with respect to $\widehat{\phi}_1$ and obtain $\widehat{\phi}_1$. Using (3.52), compute recursively $\widehat{w}_{1,t}$.

Step 4: From the second equation in (3.51), define the second sample mapping:

$$\begin{aligned}\widehat{\zeta}_1 = &\widehat{M}_T\left(\widehat{\zeta}_0\right) = \min_{\widehat{\zeta}_1} \frac{1}{T}\sum_{t=1}^{T}\left[\ln\left[\left(y_t - \widehat{\lambda}_1\widehat{\sigma}_{0,t}\right)^2\right] - \right. \\ &\left. \frac{\widehat{\mu}_{\epsilon,0}}{1-\widehat{\rho}_1} - \sum_{i=0}^{\bar{q}}\widehat{\gamma}_1\widehat{\rho}_1^i\widehat{u}_{1,t-1-i}\right]^2\end{aligned} \tag{3.54}$$

where $\widehat{\zeta}_1 = \left(\widehat{\lambda}_1, \widehat{\mu}_{\epsilon,1}, \widehat{\gamma}_1\right)'$. Similarly to Step 2 in the GARCH(1,1)-in-mean case, Proposition 1 allows to split the mapping in (3.54) such that:

$$\widehat{\lambda}_1 = \left[\widehat{\sigma}_0'\widehat{\sigma}_0\right]^{-1}\widehat{\sigma}_0' Y \tag{3.55}$$

$$\begin{aligned}\widehat{\zeta}_1^* = &\min_{\widehat{\zeta}_1^*}\frac{1}{T}\sum_{t=1}^{T}\left[\ln\left[\left(y_t - \widehat{\lambda}_1\widehat{\sigma}_{0,t}\right)^2\right] - \right. \\ &\left. \frac{\widehat{\mu}_{\epsilon,0}}{1-\widehat{\rho}_1} - \sum_{i=0}^{\bar{q}}\widehat{\gamma}_1\widehat{\rho}_1^i\widehat{u}_{1,t-1-i}\right]^2\end{aligned} \tag{3.56}$$

where $\widehat{\sigma}_0$ and Y are $(T\times 1)$ vectors stacking all observations of $\widehat{\sigma}_{0,t}$ and y_t. Compute $\widehat{\lambda}_1$ through (3.55). Plug $\widehat{\lambda}_1$ into (3.56), and minimize with respect to $\widehat{\zeta}_1^* = (\widehat{\mu}_{\epsilon,1}, \widehat{\gamma}_1)'$.

Step 5: Based on $\widehat{\phi}_1$, $\widehat{\zeta}_1$, and $\bar{\eta}^2$ solve the following system of equations to find $\widehat{\omega}_1$, $\widehat{\xi}_1$ and $\widehat{\varphi}_1$.

$$\widehat{\varphi}_1 = \frac{\widehat{\rho}_1 - \widehat{\beta}_1}{\widehat{\gamma}_1} \tag{3.57}$$

$$\widehat{\xi}_1 = \frac{\left[\widehat{\mu}_{\nu 1} - \widehat{\varphi}_1\widehat{\mu}_{\epsilon 1} + \widehat{\varphi}_1\left(1 - \widehat{\beta}_1 - \widehat{\varphi}_1\widehat{\gamma}_1\right)\bar{\eta}^2\right]}{\left(1 - \widehat{\beta}_1 - \widehat{\varphi}_1\widehat{\gamma}_1\right)} \tag{3.58}$$

$$\widehat{\omega}_1 = \widehat{\mu}_{\nu 1} - \widehat{\gamma}_1\widehat{\xi}_1 - \left(1 - \widehat{\beta}_1 - \widehat{\varphi}_1\widehat{\gamma}_1\right)\bar{\eta}^2 \tag{3.59}$$

Step 6: Recursively in (3.47) and (3.45), compute $\widehat{\sigma}^2_{1,t}$ using $\left(\widehat{\omega}_1, \widehat{\beta}_1, \widehat{\gamma}_1\right)'$. Retrieve estimates of η_t through $\widehat{\eta}_{1,t} = \frac{\left(y_t - \widehat{\lambda}_1\widehat{\sigma}_{1,t}\right)}{\widehat{\sigma}_{1,t}}$ and obtain $\widehat{\tau}_1$ by estimating (3.48) using $\widehat{\sigma}_{1,t}$ as a regressor.

Repeat Steps 3, 4, 5 and 6 until $\widehat{\theta}_j$ converges. As in the previous models, we assume convergence occurs if $\left\|\widehat{\theta}_j - \widehat{\theta}_{j-1}\right\| < c$. Note that the NL-ILS algorithm requires $\bar{\eta}^2$ to be defined exogenously, which implies that some distributional assumption has to be made on η_t. The Monte Carlo simulations showed that the NL-ILS algorithm takes more iterations to converge, which indicates that the contraction parameter associated with the RealGARCH(1,1)-in-mean model is higher than the one found in the GARCH(1,1) model. The steps above also hold for computing the RealGARCH(1,1) model. In this case, (3.55) in step 4 drops out.

3.4 Monte Carlo Study

This section addresses the performance of the NL-ILS estimator discussed in the previous section, when estimating the GARCH(1,1), weakGARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean mod-

els. We will focus on assessing consistency, efficiency and forecast performance of the NL-ILS estimator compared with the benchmark estimator: the MLE. For the GARCH(1,1)-in-mean case, we discuss an additional issue: we assess the behavior of the λ estimates (risk premium parameter) when the conditional variance is misspecified. This set of experiments plays the role of robustness analysis, since it is known that MLE estimates of λ can be biased when the conditional variance is misspecified. This follows from the fact that the information matrix is no longer block diagonal in the GARCH(1,1)-in-mean specification. In all exercises, we fix the number of replications to 1500 unless otherwise stated. We also discard the initial 500 observations to reduce dependence on initial conditions. All models are estimated using the CML[12] optimization library in GAUSS. Results for the GARCH(1,1) and weak-GARCH(1,1) models are reported in terms of the median and the relative root mean squared error (RRMSE). The relative measures are computed using the MLE benchmark (in the denominator), implying that NL-ILS outperforms the MLE estimator when the relative measures are lower than one.

The first set of simulations analyzes the performance of the NL-ILS for the GARCH(1,1) model. The data generating process follows the baseline model displayed in (3.10) and (3.11). The stochastic term η_t is set to be normally distributed, with zero mean and variance equal to one. Table 3.1 displays results for two different specifications and five different sample sizes ($T = 100$, $T = 200$, $T = 300$, $T = 400$ and $T = 500$)[13]. Despite

[12]CML (Constrained Maximum Likelihood Estimation) is library in GAUSS designed to solve maximum likelihood functions subject to linear and nonlinear constraints. In all Monte Carlo simulations, we set global variables in CML to their default values, because this specification is flexible enough to accommodate endogenous changes in both algorithms and grid search procedures.

[13]Additional results considering different specifications are available upon request.

financial data being usually available on higher frequency than macroeconomic variables, it is interesting to examine the performance of the NL-ILS and MLE estimators in small samples, following the upward bias of the GARCH parameters in the presence of structural breaks. We focus on high persistent GARCH(1,1) specifications with $(\alpha + \beta)$ close to one, because these are the most usual cases reported when modeling financial returns. As an overall picture, we find that NL-ILS estimates outperform the MLE benchmark when T is small. This was somehow expected, since MLE estimator is known to suffer from numerical problems either when T is small or $(\alpha + \beta)$ approaches to one. The outstanding performance in small samples is particularly welcome when dealing with variables that may have structural breaks and also for forecasting purposes (see the work of Giraitis, Kapetanios, and Yates (2010)). When $(\alpha + \beta) = 0.995$, the NL-ILS-ILS estimator outperforms the MLE in all sample sizes, achieving its best performance when $T = 100$, with gains of 61%, 44% and 61% for the ω, α and β parameters, respectively. Considering the specification where $(\alpha + \beta) = 0.97$, we find that the MLE estimator improves its performance, yielding more accurate estimates than the NL-ILS estimator for all sample sizes, but $T = 100$.

Table 3.2 and 3.3 report results for the weak-GARCH(1,1) model. The weak-GARCH(1,1) processes are generated in the spirit of Drost and Nijman (1993). We firstly generate a GARCH(1,1) process using (3.10) and (3.11). The vector $\theta = (\omega, \alpha, \beta)'$ used in this specification contains the high frequency parameters. The stochastic term is assumed to be normally distributed, such that $\eta_t \sim (0, 1)$. Given the high frequency GARCH(1,1) process, we re-sample y_t at different frequencies m, yielding

$y_{(m)t}$. When re-sampling, we assume y_t is a stock variable, rendering $y_{(m)t}$, $t = m, 2m, ..., T$. The low frequency parameters are computed through (3.35), (3.36) and (3.37) and denoted as θ_m. Table 3.2 displays results obtained when the high frequency α and β parameters are set equal to 0.05 and 0.94 respectively. In this scenario, as m increases (the resulting weak-GARCH(1,1) is sampled at a lower frequency), the NL-ILS estimator improves its performance when compared to the MLE benchmark. We argue that the reasons for that are twofold: firstly, as observed in the GARCH(1,1) case, NL-ILS has a better performance than MLE estimator for small samples. This plays an important role in this setup, since as m increases $T(m)$ decreases, following the fact that T (the high frequency sample size) is constant. The second reason arises from the robustness of the NL-ILS estimator to disturbances that present nonlinear dependence. Comparing the relative measures obtained in the weak-GARCH(1,1) experiment, with the ones obtained with the GARCH(1,1), we find that performance gains from the NL-ILS estimator with respect to the MLE estimator are higher for the weak-GARCH(1,1) model. It is also relevant to point out that NL-ILS improves its performance with respect to MLE estimator when β approaches one. Comparing the results when $m = 3$ in Tables 3.2 and 3.3, we find that NL-ILS improves the RRMSE in 26% and 17% for $\alpha_{(m)}$ and $\beta_{(m)}$, respectively. This is a particular relevant result, since it mimics financial series that usually display β very close to one. NL-ILS procedure delivers less biased estimates than the MLE benchmark, when the bias is assessed through the median. The reason for this discrepancy between the mean and the median arises from the presence of outliers. From the data generation process specification, α is very close to zero and β is close to

one, rendering autoregressive and moving average parameters very close to each other. This feature may lead to problems in the optimization of the sample mapping, yielding local minimums instead of global ones. This numerical problem usually generates outliers, impacting the mean and the variance of the NL-ILS estimator.

The third set of simulations focuses on analyzing the performance of the NL-ILS estimator for the GARCH(1,1)-in-mean specification. Tables 3.4 and 3.5 report results considering different specifications and sample sizes. We generate the data using (3.21), (3.22) and (3.23), where η_t is assumed to be normally distributed with zero mean and variance equal to one. To reduce the impact of the outliers when assessing the comparison between the NL-ILS and the MLE estimators, we display results in terms of relative root median squared error (RRMedSE) and the relative root median squared forecast error (RRMedSFE). The former one is adopted when analyzing the parameters of the GARCH(1,1)-in-mean, whereas the RRMedSFE is used when evaluating the out-of-sample forecast performance for both risk premium and conditional variance. We denote the out-of-sample risk premium forecast, at some horizon h, as: $\widehat{\pi}_{t+h} = \widehat{\lambda}\widehat{\sigma}_{t+h}$. As in the GARCH(1,1) and weak-GARCH(1,1) cases, the relative measures are computed having the MLE as the benchmark. We also report results considering the MLE algorithm computed using the NL-ILS estimates as starting values. We denote this results as MLE*.

Table 3.4 displays results for two different GARCH(1,1)-in-mean specifications. These specifications present $(\alpha + \beta)$ close to one (0.995 and 0.97, respectively) and $\alpha = 0.025$. With respect to the RMedSE associated with the parameters, we find that NL-ILS outperforms the MLE benchmark for

sample sizes up to $T = 300$ (except for the λ) when $(\alpha + \beta) = 0.995$. This conclusion is in line with our previous findings for the GARCH(1,1) and weak-GARCH(1,1). When T gets large, the MLE estimator performance improves reasonably fast, outperforming the NL-ILS estimator. Again, this pattern is expected, since the MLE is extremely difficult to be beaten in medium samples and when the model is correctly specified. As discussed in Section 3.2, NL-ILS is consistent, presenting a bias[14] of only 0.008 when T=750, whereas the bias associated with the MLE estimator is 0.010. The main determinant of the poorer performance of the NL-ILS estimator for large T lies on the presence of many more outliers than the ones found when the MLE algorithm is implemented. Concluding this point, we find that MLE algorithm is able to reduce the variance associated with the estimates much faster than the NL-ILS algorithm as T gets large. The poorer performance of the MLE* algorithm is also explained by the outliers. Hence, when T is small and the starting values in the MLE algorithm are very bad, the final outcome is likely to be also very poor. As expected, as T gets large MLE* algorithm converges to the standard MLE estimator.

The forecast performance of models estimated using the NL-ILS estimator is also worth highlighting. In particular, we find very good results on forecasting the conditional variance up to $T = 300$. We report gains of up to 51% in terms of the RRMedFE. The surprisingly good performance of MLE* algorithm arises from the bias on estimating ω and α. In practise, when β is very high (as is the case in this specification), models that present a bias combination such that, $\widehat{\beta}$ is downward biased and $\widehat{\omega}$ and $\widehat{\alpha}$ are upward biased, tend to perform well on forecasting, due to level effect.

[14] We compute the bias using the median within all replications.

Regarding the forecast of the risk premium, we find that NL-ILS is not able to outperform the MLE benchmark for any value of T. The reason for that is the poorer performance of the NL-ILS algorithm on estimating the risk premium parameter λ.

The second model in Table 3.4 presents a similar pattern, across different sample sizes, as the first specification discussed above. The main difference arises from the higher RRMedSE associated with the parameters (except for α). As noted before in the strong- and weak-GARCH(1,1) cases, MLE improves its performance, compared to NL-ILS, as β decreases.

Table 3.5 displays two additional GARCH(1,1)-in-mean specifications. Their main difference lies on the higher value impounded to α: $\alpha = 0.08$. Apart from the case where $T = 100$, we find that MLE delivers more accurate estimates and forecasts than the NL-ILS. This result is in line with our previous findings, indicating that the NL-ILS algorithm has an outstanding performance either when T is small or β is very close to one. Considering the forecast performance analysis, it is important to point out that, when outperformed by the MLE estimator, the NL-ILS results are, on average, only 5% to 10% worse than the results obtained with the MLE estimator.

Table 3.6 displays results considering the performance of the NL-ILS estimator when applied to the log-linear RealGARCH(1,1)-in-mean. The parameters in the RealGARCH(1,1)-in-mean are set as the ones Hansen, Huang, and Shek (2012) found in their empirical application. Overall, the results are favorable to the MLE estimator for both parameter estimation and out-of-the sample forecast of the conditional variance. The RRMSE of the parameter estimates are very high indicating a poor performance

of the NL-ILS estimator. We argue that this poor performance comes from the higher variance associated with the NL-ILS estimates (presence of outliers), since the bias assessed through the median within all replications is neglectful when T is large. In particular, it is important to point out that the MLE estimator is able to extract a considerable benefit out of the inclusion of the measurement equation. Turning the analysis to the forecast performance, the NL-ILS is able to outperform the MLE benchmark for the realized variance for sample sizes up to $T = 300$. This good performance on forecasting the realized variance contributed for a decent performance on forecasting the conditional variance.

3.4.1 Robustness

Mixed evidences, in both sign and significance of the λ parameter, have been found in the literature when estimating the risk premium using the full parametric GARCH-in-mean model and its variants. While French, Schwert, and Stambaugh (1987) found a positive value for λ, Glosten, Jagannathan, and Runkle (1993) found an opposite sign and Baillie and DeGennaro (1990) found very little evidence for a statistically significant λ. Considering the semiparametric approach, Linton and Perron (2003), Conrad and Mammen (2008) and Christensen, Dahl, and Iglesias (2012) found strong evidences of nonlinearity governing the risk premium function. Focusing on the full parametric GARCH-in-mean models, mixed results on the λ estimates can be motivated by lack of consistency of the QMLE estimator. As discussed in Bollerslev, Chou, and Kroner (1992), QMLE estimates of GARCH-in-mean parameters may be inconsistent when the conditional variance is misspecified. This drawback arises because the information ma-

trix is not block diagonal between the parameters in the conditional mean and the conditional variance. The task of correctly specifying the conditional variance is extremely difficult given the large menu of alternative models available in the literature (see Francq and Zakoian (2010), Bollerslev (2008) for a surveys on GARCH-type models). To study the performance of the NL-ILS estimator when the conditional variance is misspecified, we carry out four different experiments: in the first one, the conditional variance is specified as being an APARCH model (see Ding, Granger, and Engle (1993)); in the second exercise, the conditional variance is set to follow an EGARCH model in the spirit of Nelson (1991); the third exercise consists on modeling the conditional variance as a JGR-GARCH as in Glosten, Jagannathan, and Runkle (1993); the final simulation is carried out using a GARCH(2,2) specification. Note that both EGARCH and JGR-GARCH models capture asymmetric responses of the conditional variance to positive and negative shocks, whereas the APARCH specification manages to capture three important stylized facts: long memory, dependence on some power transformation of the conditional standard deviations and asymmetric responses to positive and negative shocks. We assess performance on estimating λ through the RRMSE and bias. Forecast performance is assessed using the RRMSFE. As in the previous experiments, the MLE estimator is the benchmark for all the relative measures. We also report the results for the MLE estimates which are computed using the NL-ILS estimates as starting values (MLE*).

Table 3.7 displays results for the APARCH and EGARCH models. Considering the APARCH results, we find that the NL-ILS estimator outperforms the MLE benchmark for all the different sample sizes. The difference

in performance achieves 22% when $T = 1750$. When looking at the bias, the conclusion is even more favorable to the NL-ILS estimator. We find that MLE estimates are downward biased in 15%, when $T = 1750$, whereas the bias related to the NL-ILS is neglectful. In spite of the good performance on estimating λ, the NL-ILS algorithm fails on achieving outstanding results on forecasting both the risk premium and the conditional variance. In both cases, the MLE estimator delivers more accurate forecasts.

When the conditional variance is misspecified using the EGARCH specification, the results regarding the estimation of λ are, again, extremely favorable to the NL-ILS estimator. Table 3.7 reports gains of 47% in terms of the RRMSE with respect to the MLE benchmark. The outstanding difference in performance is consistent through all the sample sizes, showing the robustness of the NL-ILS estimator. Analyzing the bias computed from both estimators, we find a similar picture as in the APARCH case: NL-ILS delivers neglectful bias, indicating consistency, whereas MLE is upward biased in 15%. A different picture arises when considering the forecast performance of the risk premium function. The NL-ILS estimator is now able to outperform the MLE benchmark in up to 28%, considering the median within all forecast horizons. We claim that this difference in performance comes mostly from the best estimation of λ, since NL-ILS does a worse job on forecasting the conditional variance.

Table 3.8 displays results for models using GJR-GARCH and GARCH(2,2) specifications. Considering the latter one, the results are very similar to the standard GARCH(1,1)-in-mean experiments carried out previously in this section. Overall, MLE provides more accurate results for both the estimation of λ and the variance and risk premium forecasts. It is important

to point out, however, that bias associated with the NL-ILS estimates of λ is neglectful. The results associated with the GJR-GARCH model follows the same pattern as the ones obtained with the APARCH and EGARCH specifications. We find that the MLE estimates of λ are biased, leading to a poorer performance of this estimator when compared to the NL-ILS procedure.

Overall, the conclusion obtained from this set of experiments is that NL-ILS is more robust than the MLE benchmark when the conditional variance is misspecified. Moreover, MLE delivers biased estimates of λ when the conditional variance is misspecified in such a way that it possesses either asymmetric responses to positive and negative shocks or dependence at different moments than the second one. Finally, we claim that, under misspecification of the conditional variance, inference using the MLE framework may no longer be a valid alternative.

3.5 Empirical application

We examine the significance of the risk premium parameter using the GARCH(1,1)-in-mean framework by adopting the NL-ILS estimator discussed in the previous sections. As discussed in the Monte Carlo section, the performance of both NL-ILS and QMLE estimator may vary when dealing with weak processes, such as the weak-GARCH(1,1). As the true data generation process governing the excess returns are believed not to be discrete (such as daily, weekly or monthly), the impact of time aggregation on the consistency of the λ estimates needs to be addressed. To this purpose, we construct nine different data sets on excess returns, compre-

hending three different indices (CRSP value-weighted index, S&P500 and S&P100) at three different frequencies: daily, weekly and monthly. The CRSP value-weighted index is considered as the most complete (in market sense) index, being therefore the best *proxy* for the market, as pointed out by Linton and Perron (2003). Hence, by using "least complete" indices, we check whether the significance of the risk premium parameter depends on the market coverage of the index. Excess returns for the three indices are computed deducting the risk free rate (one-month Treasury bill rate) from their log returns[15]. Table 3.9 reports the descriptive statistics. The daily CRSP and S&P500 indices spans from 28/06/1963 to 29/09/2011, accounting for 12,148 observations. The S&P100 index spans from a smaller period, (04/08/1982 - 29/09/2011), yielding 7,364 observations. CRSP and S&P500 indices have 2,426 and 740 observations for weekly and monthly frequencies, respectively. S&P100 index contains 1,469 and 330 observations on the weekly and monthly frequencies, respectively. Standard errors for the NL-ILS estimator are computed using block bootstrap with one thousand replications, whereas for the QMLE estimator the Bollerslev-Wooldridge robust standard errors are implemented.

We start discussing the results reported in Table 3.10, where we estimate a GARCH(1,1)-in-mean model using both, NL-ILS and QMLE estimators. At daily frequency, the λ estimates obtained using the NL-ILS estimator are significant at 5%[16] level only for the CRSP index, whereas QMLE estimator delivers significant estimates for all series. Considering the parameters

[15]CRSP value-weighted index and one-month Treasure bill rate were downloaded from WRDS - Wharton Research Data Services, whereas S&P500 and S&P100 indices were obtained from Yahoo! finance.

[16]T-statistics for the NL-ILS estimate of λ is on the boundary of 5% level significance. However, looking at the empirical distribution computed from the bootstrapped estimates of λ, the NL-ILS turned out to be significant at 5% level.

in the conditional variance equation, both procedures deliver highly significant estimates of the parameters α and β. Both methodologies also yield high degrees of persistence ($\alpha+\beta$). With respect to this point, it is relevant to mention that, in all indices, the persistence obtained using QMLE estimator is always higher (average of 0.99502) than the ones obtained using the NL-ILS estimator. To check this issue, we also performed the MLE estimation using the GED distribution. Results did not show any significant quantitative change. There is an important difference in magnitude from the λ estimates obtained with the NL-ILS algorithm and the ones obtained with the MLE methodology. In fact, the difference between them turned out to be statistically significant[17], indicating the possibility of MLE being upward biased following discussion in Section 3.4.

Moving to the weekly and monthly frequencies, their patterns remain very similar to the one previously discussed. NL-ILS delivers λ estimates which are significant at 5% level only for the CRSP index, whereas QMLE estimates are significant for all indices. The α and β parameters from the conditional variance equation remain highly significant, yielding a high degree of persistence in the conditional variance. The results in Table 3.10 turn out to be consistent with the previous findings in the literature. Linton and Perron (2003) found a value of $\widehat{\lambda}$ when estimating a EGARCH-in-mean very close to the NL-ILS estimates. The same applies for Christensen, Dahl, and Iglesias (2012), who found significant QMLE estimates of λ for the daily S&P500 index.

As a second step of our investigation, we incorporate realized measures of volatility in this analysis by estimating the RealGARCH(1,1)-in-mean

[17]T-statistics are 4.77, 3.44, 2.50 for CRSP, S&P500 and S&P100, respectively.

for the S&P500 index[18]. Results in Table 3.11 corroborate our previous findings: risk premium parameter is not significant to any of the sample frequencies.

Analyzing the empirical results at the light of the results obtained in the Monte Carlo section, we claim that this difference in magnitude and significance may be caused by bias on the QMLE estimates, following misspecification of the conditional variance. Assuming our claim is correct, we outline two conclusions: firstly, the risk premium parameter is only significant for the most complete index (CRSP), whereas for the "less complete" indices the risk-return tradeoff does not hold. This finding is consistent with the theoretical results in Merton (1973), that requires the existence of a market portfolio. Hence, the fact that the λ estimates obtained from the S&P500 and S&P100 are not significant may imply that these two indices are not good *proxies* for the market. Secondly, we conclude that the NL-ILS estimator is the most suitable for dealing with the task of estimating the risk premium parameter, since, as observed in the Section 3.4, it is robust to misspecification of the conditional variance.

3.5.1 Empirical application: Robustness

As a robustness check, we estimate the risk premium parameter using three alternative models: APARCH(1,1,1)-in-mean, EGARCH(1,1,1)-in-mean and GJR-GARCH(1,1,1)-in-mean models. All the three models are estimated using the QMLE procedure. Table 3.12 reports results for all indices at all frequencies. By using models that allow for asymmetric re-

[18]Realized measures of the conditional variance were obtained from the Oxford-Man Institute of Quantitative Finance (realized Library). Unfortunately, among the three different indices we adopt in this chapter, there is only availability of data for the S&P500 index.

sponse of the conditional variance to positive and negative shocks, it turns out that the risk premium parameter λ is only significant for the CRSP index. This finding strengths the conclusion that the GARCH(1,1)-in-mean estimated with the MLE estimator is not robust enough to misspecification of the conditional variance equation, leading to misleading results.

3.6 Conclusion

This chapter introduces a novel estimator: the nonlinear iterative least squares (NL-ILS). To illustrate the NL-ILS estimator, we provide algorithms covering the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. We show that the NL-ILS estimator is particulary useful when innovations in the mean equation have some degree of dependence or the variance equation is misspecified. These both features may lead to inconsistency when the QMLE procedure is implemented. We establish the consistency and asymptotic distribution for the NL-ILS estimator covering the GARCH(1,1) model and extend the consistency result for the weak-GARCH(1,1) model. The assumptions we require for the asymptotic theory are compatible with the QMLE estimator. Through an extensive Monte Carlo study, we show that the NL-ILS estimator outperforms the MLE benchmark in a variety of scenarios including the following: the sample size is small; the β parameter in the conditional variance has values very close to one, as widely found in empirical studies; or the true data generation process (DGP) is the weak-GARCH(1,1), indicating that the NL-ILS estimator is more robust to the presence of dependence on the in-

novations. Moreover, we show that the NL-ILS estimator is more robust to misspecification of the conditional variance, delivering neglectful biases on estimating the risk premium parameter in a GARCH(1,1)-in-mean model. In contrast with the NL-ILS algorithm, the MLE estimator presents biases of up to 15%, leading to the differences in performances of up to 22%, in terms of the relative mean squared error, when estimating the risk premium parameter. The NL-ILS estimator also delivers more accurate out-of-the-sample forecasts for the risk premium function when the DGP is either the EGARCH(1,1,1)-in-mean or the GJR-GARCH(1,1,1)-in-mean models.

An empirical application addressing the significance of the risk premium parameter through a full parametric GARCH-in-mean and RealGARCH (1,1)-in-mean models is provided. We undertake our analysis through two different dimensions: temporal aggregation and market representation. The latter dimension is appraised by using the CRSP, S&P500 and S&P100 indices, which possess distinct market coverage, whereas the former dimension is assessed by aggregating the series at daily, weekly and monthly basis. When adopting the robust NL-ILS estimator and the QMLE benchmark to assess significance of the risk premium parameter, the results turned out to be very different: the NL-ILS estimator delivered risk premium estimates which are significant only for the CRSP index at all its frequencies; the QMLE estimator, however, provides estimates which are significant to all three data sets, in all frequencies. Moreover, the difference in magnitude between the NL-ILS and QMLE estimates are also significant in some data sets, indicating a potential bias. By using the Monte Carlo results, we argue that the QMLE estimator provides biased estimates following a misspecified conditional variance. As a robustness check for the empir-

ical results, we estimate RealGARCH(1,1)-in-mean, EGARCH(1,1,1)-in-mean, APARCH(1,1,1)-in-mean and GJR-GARCH(1,1,1)-in-mean models and their results corroborate our findings using the GARCH(1,1)-in-mean estimated with NL-ILS algorithm: the risk premium parameter is only significant for the CRSP index at all frequencies. Ultimately, this chapter suggests the use of the NL-ILS estimator on modeling the conditional volatility in the presence of dependent errors and misspecification. We highlight the robustness properties of the NL-ILS estimator assessing the risk premium in different indices and sampling frequencies.

3.7 Appendix

Proof of Lemma 7: It mirrors the proof in Lemma 5 in Dominitz and Sherman (2005) and Chapter 2. Define the population mapping evaluated at some vectors of parameters ξ and ς , such that $\xi, \varsigma \in \mathbb{B}$. By Taylor expansion, rewrite $|N(\xi) - N(\varsigma)|$ defining a bound that contains the gradient of the population mapping evaluated on ϕ.

$$\begin{aligned} &|N(\xi) - N(\varsigma)| = |V(\xi^*)[\xi - \varsigma]| \\ &|V(\xi^*)[\xi - \varsigma]| \leq |V(\phi)[\xi - \varsigma]| + |[V(\xi^*) - V(\phi)][\xi - \varsigma]| + \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad o_p(|\xi - \varsigma|) \end{aligned} \tag{3.60}$$

Using Dominitz and Sherman (2005) result (Lemma 5), it suffices to show that the maximum eigenvalue of $V(\phi)$ is less than one in absolute value to prove Lemma 7. By applying the NR procedure, the population and sample mapping in (3.14) and (3.15) can be linearized as:

$$\phi_{j+1} = N(\phi_j) = \phi_j - [H(\phi_j)]^{-1} G(\phi_j) \tag{3.61}$$

$$\widehat{\phi}_{j+1} = \widehat{N}_T\left(\widehat{\phi}_j\right) = \widehat{\phi}_j - \left[\widehat{H}_T\left(\widehat{\phi}_j\right)\right]^{-1} \widehat{G}_T\left(\widehat{\phi}_j\right) \tag{3.62}$$

where $\widehat{G}_T\left(\widehat{\phi}_j\right)$ and $\widehat{H}_T\left(\widehat{\phi}_j\right)$ are the gradient and Hessian of $Q_T\left(\widehat{\phi}_{j+1}\right)$ evaluated on $\widehat{\phi}_j$, and $G(\phi_j)$ and $H(\phi_j)$ are their population counterparts. Using (3.61), the gradient of the population mapping on the $(j+1)^{th}$ iter-

ation, defined as $V(\phi_j) = \bigtriangledown_{\phi_j} N(\phi_j)$, is given by:

$$V(\phi_j)\Big|_{\phi} = \left[\bigtriangledown_{\phi_j} N(\phi_j)\right]\Big|_{\phi} = I_3 - \left\{\left[I_1 \otimes \left[H(\phi_j)\right]^{-1}\Big|_{\phi}\right] \times \right. \tag{3.63}$$
$$\left. \frac{\partial vec\left(G(\phi_j)\right)}{\partial \phi_{\prime}}\Big|_{\phi} + \left[G(\phi_j)' \otimes I_3\right]\Big|_{\phi} \frac{\partial vec\left(\left[H(\phi_j)\right]^{-1}\right)}{\partial \phi'}\Big|_{\phi}\right\}$$

When evaluated at the true vector of parameters, the second term on the right-hand side of (3.63) is zero, following $\left[G(\phi_j)' \otimes I_3\right]\Big|_{\phi} = 0$. Hence, (3.63) reduces to:

$$V(\phi_j)\Big|_{\phi} = I - \left[H(\phi_j)\right]^{-1}\Big|_{\phi}\left[\bigtriangledown_{\phi_j} G(\phi_j)\right]\Big|_{\phi} \tag{3.64}$$

The expressions for $\left[H(\phi_j)\right]^{-1}\Big|_{\phi}$ and $\left[\bigtriangledown_{\phi_j} G(\phi_j)\right]\Big|_{\phi}$ are given by:

$$\left[H(\phi_j)\right]^{-1}\Big|_{\phi} = \begin{pmatrix} \frac{(-1+a)}{2}\left[\frac{-1+a-((1+a)^3\omega^2)}{((a+b)^2\sigma_u^2)}\right] & -\frac{(-1+a)^2(1+a)^3\omega}{2(a+b)^2\sigma_u^2} & \\ -\frac{(-1+a)^2(1+a)^3\omega}{2(a+b)^2\sigma_u^2} & -\frac{(-1+a^2)^3}{2(a+b)^2\sigma_u^2} & \vdots \\ -\frac{(-1+a)(1+a)^2(1+ab)\omega}{2(a+b)^2\sigma_u^2} & -\frac{(-1+a^2)^2(1+ab)}{2(a+b)^2\sigma_u^2} & \end{pmatrix} \tag{3.65}$$
$$\begin{pmatrix} -\frac{(-1+a)(1+a)^2(1+ab)\omega}{2(a+b)^2\sigma_u^2} \\ -\frac{(-1+a^2)^2(1+ab)}{2(a+b)^2\sigma_u^2} \\ -\frac{(-1+a^2)(1+4ab+b^2+a^2(1+b^2))}{2(a+b)^2\sigma_u^2} \end{pmatrix}$$

$$\left[\bigtriangledown_{\phi_j} G\left(\phi_j\right)\right]\Big|_{\phi} = \begin{pmatrix} -\frac{2}{(-1+a)(1+b)} & \frac{2\omega}{(-1+a)^2(1+b)} & 0 \\ \frac{2\omega}{(-1+a)^2(1+b)} & V_{22} & V_{23} \\ 0 & -\frac{2(1+a+b)(1+ab)\sigma_u^2}{(-1+a)(1+a)^2(1+b)} & \frac{2(1+a+b)\sigma_u^2}{(1+a)(1+b)} \end{pmatrix} \tag{3.66}$$

with $V_{22} = \frac{-2(1+a)^3\omega^2 - 2(1-a^2+b+a(1+a-4a^2+a^4)b+a(4-5a+a^3)b^2+(-1+a)^2b^3)\sigma_u^2}{(-1+a^2)^3(1+b)}$ and $V_{23} = \frac{2((-1+a)(1+a)^2+(-1+a+a^2)b+b^2)\sigma_u^2}{(-1+a)(1+a)^2(1+b)}$.

Using results in (3.66) and (3.65) and collecting terms in (3.64), $V\left(\phi\right)$ reduces to:

$$V\left(\phi\right) = \begin{pmatrix} \frac{a+b}{1+b} & -\frac{(1+(-1+a+a^2)b)\omega}{1+b} & \frac{(-1+a)(1+a)^2\omega}{1+b} \\ 0 & \frac{(1+2ab-a^3b)}{1+b} & \frac{(-1+a^2)^2}{1+b} \\ 0 & -\frac{(1+ab)^2}{1+b} & \frac{(-1+a^2(2+ab))}{1+b} \end{pmatrix} \tag{3.67}$$

Define the Eigenvalues associated with $V\left(\phi\right)$ as $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)'$. By solving (3.67), ε is given by:

$$\varepsilon = \left[\frac{a+b}{1+b}, \quad \frac{a(a+b)}{1+b}, \quad \frac{a(a+b)}{1+b}\right]' \tag{3.68}$$

Remark: In Lemma 7, it is important to point out that the eigenvalues associated with (3.67) do not depend on ω nor on σ_u^2. This allows to focus only with the parameters a, b which are bounded by Assumption B1. To this purpose, we evaluate the properties of (3.68) performing a numeri-

cal grid search through different combinations of parameters a and b that satisfy Assumption B1. Figure 3.1 displays the maximum eigenvalue computed using (3.68). From Figure 3.1, the maximum eigenvalue associated with $V(\phi)$ is smaller than one, in absolute value, for all combinations in the grid search. This is enough to prove Lemma 7, yielding that $N(\xi)$ is an ACM for all $\xi \in \mathbb{B}$.

Lemma 8 *Denote* $\Psi_i = \varrho_i(\phi)$ *and* $\widetilde{\Psi}_i = \varrho_i\left(\widetilde{\phi}\right)$, *with* ϕ, $\widetilde{\phi} \in \mathbb{B}$. *Suppose Assumptions B1, B2 and B3 hold. Then,*

$$\mathbb{E}\left\{\left[\left(\psi_{j+1,0} - \sum_{i=0}^{\infty}\psi_{j+1,i}u_{j,t-1-i}\right) - \left(\widetilde{\psi}_{j+1,0} - \sum_{i=0}^{\infty}\widetilde{\psi}_{j+1,i}u_{j,t-1-i}\right)\right]^2\right\} \neq 0$$

for all $\widetilde{\phi} \neq \phi$

Proof of Lemma 8: To prove Lemma 8, rewrite the target expression as:

$$\mathbb{E}\left\{\left[\left(\psi_0 - \widetilde{\psi}_0\right) - \left(\sum_{i=0}^{\infty}\widetilde{\psi}_i u_{t-1-i} - \sum_{i=0}^{\infty}\psi_i u_{t-1-i}\right)\right]^2\right\} \neq 0$$

$$\mathbb{E}\left\{\left(\psi_0 - \widetilde{\psi}_0\right)^2 - 2\left(\psi_0 - \widetilde{\psi}_0\right)\sum_{i=0}^{\infty}\left(\widetilde{\psi}_i - \psi_i\right)u_{t-1-i} + \left(\sum_{i=0}^{\infty}\widetilde{\psi}_i u_{t-1-i} - \sum_{i=0}^{\infty}\psi_i u_{t-1-i}\right)^2\right\} \neq 0$$

$$\left(\psi_0 - \widetilde{\psi}_0\right)^2 + \mathbb{E}\left\{\left[\sum_{i=0}^{\infty}\left(\widetilde{\psi}_i - \psi_i\right)u_{t-1-i}\right]^2\right\} \neq 0$$

$$\left(\psi_0 - \widetilde{\psi}_0\right)^2 + \mathbb{E}\left\{\sum_{i=0}^{\infty}\left(\widetilde{\psi}_i - \psi_i\right)^2 u_{t-1-i}^2 + 2\sum_{i=0}^{\infty}\left(\widetilde{\psi}_i - \psi_i\right)u_{t-1-i}\left[\sum_{l=i+1}^{\infty}\left(\widetilde{\psi}_l - \psi_l\right)u_{t-1-l}\right]\right\} \neq 0$$

$$\left(\psi_0 - \widetilde{\psi}_0\right)^2 + \sum_{i=0}^{\infty}\left(\widetilde{\psi}_i - \psi_i\right)^2\sigma_u^2 \neq 0 \tag{3.69}$$

Showing that (3.69) holds is equivalent to prove:

$$\left|\psi_0 - \widetilde{\psi}_0\right| + \sigma_u^2 \sum_{i=0}^{\infty} \left|\widetilde{\psi}_i - \psi_i\right| > 0 \tag{3.70}$$

We shall prove (3.70) by contradiction. To this purpose, we show that $\widetilde{\phi} = \phi$ is the only vector that sets (3.70) to zero. Define ϕ^* as vector located in the segment line between ϕ and $\widetilde{\phi}$. Using the first order Taylor expansion, the first term on the left-hand side of (3.70) reduces to:

$$\left|\psi_0 - \widetilde{\psi}_0\right| = \left|\frac{\partial\left(\frac{\omega}{1-a}\right)}{\partial\phi'}\bigg|_{\phi^*}\right| \left|\phi - \widetilde{\phi}\right| = \left|\left[\frac{1}{1-a^*},\ \frac{\omega^*}{(1-a^*)^2},\ 0\right]\right| \left|\phi - \widetilde{\phi}\right| \tag{3.71}$$

Assumption B1 guarantees that the first two elements of $\frac{\partial\psi_0}{\partial\phi'}\Big|_{\phi^*}$ are strictly positive. Given that, (3.71) is equal to zero only if $\widetilde{\phi} = \left[\omega, a, \widetilde{b}\right]'$, for any $\widetilde{b}$ satisfying Assumption B1. Hence it makes necessary to show that the second term on the left-hand side of (3.70) is greater than zero when evaluated at $\widetilde{\phi} = \left[\omega,\ a,\ \widetilde{b}\right]'$. To this purpose, we apply the first order Taylor expansion such that:

$$\begin{aligned} \sigma_u^2 \sum_{i=0}^{\infty} \left|\widetilde{\psi}_i - \psi_i\right| &= \sigma_u^2 \sum_{i=0}^{\infty} \left|\frac{\partial\psi_i}{\partial\phi}\bigg|_{\phi^*}\right| \left|\phi - \widetilde{\phi}\right| \\ &= \sigma_u^2 \sum_{i=0}^{\infty} \left|\left[0,\ ia^*\left(a^* + b^*\right),\ a^{*i}\right]\right| \left|\phi - \widetilde{\phi}\right| \end{aligned} \tag{3.72}$$

The third element of $\frac{\partial\psi_i}{\partial\phi}\Big|_{\phi^*}$ is strictly greater than zero for all $i \geq 1$ and b^* satisfying Assumption B1, implying that when evaluated on $\widetilde{\phi} = \left[\omega,\ a,\ \widetilde{b}\right]'$ (3.72) is strictly greater than zero. Hence, the only vector that sets (3.70) to zero is $\widetilde{\phi} = \phi$. This concludes the proof of Lemma 8.

Lemma 9 *Suppose Assumptions B1, B2 and B3 hold. Then,*

$\sup_{\xi\in\mathbb{B}}\left|\widehat{N}_T(\xi)-N(\xi)\right|=o_p(1)$ *as* $T\longrightarrow\infty$

Proof of Lemma 9: By evaluating both (3.61) and (3.62) on ϕ_j, the absolute difference between the population mapping and its sample counterpart is given by:

$$\left|\widehat{N}_T(\phi_j)-N(\phi_j)\right|=\left|\left[\phi_j-[H(\phi_j)]^{-1}G(\phi_j)\right]-\left[\phi_j-\left[\widehat{H}_T(\phi_j)\right]^{-1}\widehat{G}_T(\phi_j)\right]\right| \tag{3.73}$$

Subtracting and adding $\left[\widehat{H}_T(\phi_j)\right]^{-1}G(\phi_j)$ in (3.73):

$$\left|\widehat{N}_T(\phi_j)-N(\phi_j)\right|\leq\left|\left[\left[\widehat{H}_T(\phi_j)\right]^{-1}-[H(\phi_j)]^{-1}\right]G(\phi_j)\right|-\left|\left[\widehat{H}_T(\phi_j)\right]^{-1}\left[\widehat{G}_T(\phi_j)-G(\phi_j)\right]\right| \tag{3.74}$$

To prove point-wise convergence of the population and sample mappings evaluated at the same vector of parameters, it suffices to show that both terms on the right-hand side of (3.74) have order $o_p(1)$ as $T\longrightarrow\infty$. This implies showing that sample gradient and Hessian converge to their population counterparts, when evaluated on the true vector of parameters ϕ.

The sample and population gradient are then given by:

$$\widehat{G}_T(\phi) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{-2}{1-a}\dot{u}_t\right\} \\ \frac{1}{T}\sum_{t=1}^{T}\left\{2\dot{u}_t\left(-\frac{\omega}{(1-a)^2}-\sum_{i=0}^{\bar{q}}d_i u_{t-1-i}\right)\right\} \\ \frac{1}{T}\sum_{t=1}^{T}\left\{-2\left(\sum_{i=0}^{\bar{q}}a^i u_{t-1-i}\right)\dot{u}_t\right\} \end{pmatrix} \tag{3.75}$$

$$G(\phi) = \begin{pmatrix} \mathbb{E}\left\{\frac{-2}{1-a}\left(\dot{u}_t - \sum_{i=\bar{q}+1}^{\infty}a^i(a+b)u_{t-1-i}\right)\right\} \\ \mathbb{E}\left\{2\left(\dot{u}_t - \sum_{i=\bar{q}+1}^{\infty}a^i(a+b)u_{t-1-i}\right)\times \right. \\ \qquad \left.\left(-\frac{\omega}{(1-a)^2}-\sum_{i=0}^{\bar{q}}d_i u_{t-1-i}-\sum_{i=\bar{q}+1}^{\infty}d_i u_{t-1-i}\right)\right\} \\ \mathbb{E}\left\{-2\left(\sum_{i=0}^{\bar{q}}a^i u_{t-1-i}-\sum_{i=\bar{q}+1}^{\infty}a^i u_{t-1-i}\right)\dot{u}_t\right\} \end{pmatrix} \tag{3.76}$$

where $d_i = a^i + ia^{i-1}(a+b)$ and $\dot{u}_t = \epsilon_t^2 - \frac{\omega}{1-a} - \sum_{i=0}^{\bar{q}} a^i(a+b)u_{t-1-i}$. Provided that $\bar{q} \longrightarrow \infty$ as $T \longrightarrow \infty$, and $\sum_{i=0}^{\infty}|a^i| < \infty$ following $|a| < 1$, the additional terms in the population mapping converges in probability to zero as $T \longrightarrow \infty$, such that $\sum_{i=\bar{q}+1}^{\infty} a^i \xrightarrow{p} 0$ and $\sum_{i=\bar{q}+1}^{\infty} d_i \xrightarrow{p} 0$. Furthermore, all elements in (3.75) are averages of m.d.s. processes. This allows the use of the weak law of large numbers, such that (3.77) holds. Similar steps are conducted to show that the sample Hessian converges in probability to its population counterpart, yielding (3.78).

$$\widehat{G}_T(\phi_j) \xrightarrow{p} G(\phi_j) \tag{3.77}$$

$$\widehat{H}_T(\phi_j) \xrightarrow{p} H(\phi_j) \tag{3.78}$$

To obtain uniform convergence in probability, the sample mapping needs

to be stochastically equicontinuous. Assumption B3 provides the Lipschitz condition for $\widehat{N}_T(\phi_j)$ for all $\phi_j \in \mathbb{B}$. Following Lemma 2.9 in Newey and McFadden (1994), the Lipschitz condition implies that the sample mapping is stochastically equicontinuous, which allows the use of theorem 21.9 (pg. 337) in Davidson (1994), yielding uniform convergence between sample and population mappings.

Lemma 10 *Suppose Assumptions B1, B2 and B3 hold and fix* ξ *and* ς *in* $\mathbb{B}$. *Then,*

$\sup_{\xi,\varsigma\in\mathbb{B}} \left|\widehat{\Lambda}_T(\xi,\varsigma) - \Lambda(\xi,\varsigma)\right| = o_p(1)$ *as* $T \longrightarrow \infty$

Proof of Lemma 10: Proof of Lemma 10 mirrors the steps of Lemma 4 in Chapter 2. Using their result, rewrite $\sup_{\xi,\varsigma\in\mathbb{B}} \left|\widehat{\Lambda}_T(\xi,\varsigma) - \Lambda(\xi,\varsigma)\right| = o_p(1)$ as:

$$\sup_{\xi,\varsigma\in\mathbb{B}} \left|\Lambda(\xi,\varsigma) - \widehat{\Lambda}_T(\xi,\varsigma)\right| \leq \frac{1}{|\xi-\varsigma|}\left[\sup_{\xi,\varsigma\in\mathbb{B}} \left|N(\xi) - \widehat{N}_T(\xi)\right| + \sup_{\xi,\varsigma\in\mathbb{B}} \left|N(\varsigma) - \widehat{N}_T(\varsigma)\right|\right] \tag{3.79}$$

Lemma 9 implies that both terms inside the brackets have order $o_p(1)$. Assumption B1 states that $[\xi - \varsigma]$ is bounded, implying that the right-hand side of (3.79) converges in probability to zero, as $T \longrightarrow \infty$.

Lemma 11 *Suppose Assumptions B1, B2 and B3 hold and fix* ξ *and* ς *in* $\mathbb{B}$. *If*

i) $\sup_{\xi\in\mathbb{B}} \left|\widehat{N}_T(\xi) - N(\xi)\right| = o_p(1)$ *as* $T \longrightarrow \infty$

ii) $\sup_{\xi,\varsigma\in\mathbb{B}} \left|\widehat{\Lambda}_T(\xi,\varsigma) - \Lambda(\xi,\varsigma)\right| = o_p(1)$ *as* $T \longrightarrow \infty$

then, $\widehat{N}_T(\xi)$ *is an ACM on* $(\mathbb{B}, d)$, *with* $\xi \in \mathbb{B}$ *and it has fixed point denoted by* $\widehat{\phi}$, *such that* $\left|\widehat{\phi}_j - \widehat{\phi}\right| = o_p(1)$, *as* $j \longrightarrow \infty$ *with* $T \longrightarrow \infty$.

Proof of Lemma 11: see Lemma 5 in Chapter 2.

Lemma 12 *Suppose Assumptions B1, B2 and B3 hold. If* $\widehat{N}_T(\gamma)$ *is an ACM on* $(\mathbb{B}, d)$
Then, $\sqrt{T}\left|\widehat{\phi}_j - \widehat{\phi}\right| = o_p(1)$ *as* $T \longrightarrow \infty$ *and* $j \longrightarrow \infty$

Proof of Lemma 12: Lemma 6 in Chapter 2 gives:

$$\sqrt{T}\left|\widehat{\phi}_j - \widehat{\phi}\right| \leq \sqrt{T}\widehat{\kappa}^j \left|\widehat{\phi}_0 - \widehat{\phi}\right| \tag{3.80}$$

The right-hand side converges in probability to zero if $\frac{\ln(T)}{j} = o(1)$. In fact, $j \gg -\frac{1}{2}\left[\frac{\ln(T)}{\ln(\kappa)}\right]$ needs to hold, implying that speed of convergence depends on the contraction parameter of the population mapping.

Proof of Theorem 2: We divide this proof in two sections. In the first part, we prove the consistency of the NL-ILS estimator (item (i) in Theorem 2), whereas the second part focuses on the asymptotic distribution (part (ii) in Theorem 2). From Dominitz and Sherman (2005), if $N(\xi)$ is an ACM on $(\mathbb{B}, d)$, then $N(\xi)$ is also a contraction map. Lemmas 7 and 11 state that the population and the sample mapping are ACM. These allow the use of standard fixed-point theorem as stated in Burden and Faires (1993) and Judd (1998) to show consistency of the NL-ILS estimator. Identification on the population mapping gives $N(\phi) = \phi$. To show that $\left|\widehat{\phi} - \phi\right| = o_p(1)$, rewrite this term

$$\left|\widehat{\phi} - \phi\right| \leq |\phi_j - \phi| + \left|\widehat{\phi} - \phi_j\right| \tag{3.81}$$

The first term on the right-hand side can be expressed only as function of the population mapping. Rewriting it in this way and substituting recur-

sively using the ACM bound, $|\phi_j - \phi|$ resumes to

$$
\begin{aligned}
&|\phi_j - \phi| = |N(\phi_{j-1}) - N(\phi)| \leq \kappa |\phi_{j-1} - \phi| \\
&|\phi_j - \phi| \leq \kappa |N(\phi_{j-1}) - N(\phi)| \leq \kappa^2 |\phi_{j-1} - \phi| \\
&|\phi_j - \phi| \leq \kappa^j |N(\phi_0) - N(\phi)|
\end{aligned} \tag{3.82}
$$

Provided that $j \longrightarrow \infty$ as $T \longrightarrow \infty$, the right-hand side of (3.82) converges in probability to zero. Hence, to show consistency of the NL-ILS estimator it remains to show that the second term on the right-hand side of (3.81) converges in probability to zero. Rewrite this term as:

$$
\left|\widehat{\phi} - \phi_j\right| \leq \left|\widehat{\phi} - \widehat{\phi}_j\right| + \left|\widehat{\phi}_j - \phi_j\right| \tag{3.83}
$$

The first term on the right-hand side of (3.83) has order $o_p\left(T^{\frac{1}{2}}\right)$ following Lemma 12. The second term on the right-hand side of (3.83) is bounded as

$$
\left|\widehat{\phi}_j - \phi_j\right| \leq \left|\widehat{N}_T\left(\widehat{\phi}_{j-1}\right) - N\left(\widehat{\phi}_{j-1}\right)\right| + \left|N\left(\widehat{\phi}_{j-1}\right) - N(\phi_{j-1})\right| \tag{3.84}
$$

The first term on the right-hand side of (3.84) has order $o_p(1)$ following Lemma 9. The remaining term of (3.84) can be rewritten using the ACM bound, such that:

$$
\left|N\left(\widehat{\phi}_{j-1}\right) - N(\phi_{j-1})\right| \leq \kappa \left|\widehat{\phi}_{j-1} - \phi_{j-1}\right| \tag{3.85}
$$

Applying recursively the same strategy as in (3.84) and (3.85), equation

(3.83) reduces to

$$\left|\widehat{\phi}-\phi_j\right| \leq \kappa^j \left|\widehat{\phi}_0-\widehat{\phi}_0\right| \tag{3.86}$$

Note that $\left|\widehat{\phi}_0-\phi_0\right|$ is bounded, provided that $\widehat{\phi}_0,\ \phi_0 \in \mathbb{B}$. As $j \longrightarrow \infty$ with $T \longrightarrow \infty$, the right-hand side of (3.86) has order $o_p(1)$, implying $\left|\widehat{\phi}-\phi\right| = o_p(1)$.

We now prove the asymptotic distribution of the NL-ILS estimator. This proof mirrors the steps of Theorem 4 in Dominitz and Sherman (2005). To establish the asymptotic distribution of$\sqrt{T}\left[\phi_j-\phi\right]$, firstly rewrite it as:

$$\sqrt{T}\left[\widehat{\phi}_j-\phi\right] = \sqrt{T}\left[\widehat{\phi}_j-\widehat{\phi}\right] + \sqrt{T}\left[\widehat{\phi}-\phi\right] \tag{3.87}$$

The first term on the right-hand side of equation (3.87) has order $o_p(1)$ following Lemma 12 and provided that $\frac{\ln(T)}{j} = o(1)$. The second term of (3.87) resumes to

$$\begin{aligned}\sqrt{T}\left[\widehat{\phi}-\phi\right] &= \sqrt{T}\left[\widehat{N}_T\left(\widehat{\phi}\right)-N(\phi)\right] \\ \sqrt{T}\left[\widehat{N}_T\left(\widehat{\phi}\right)-N(\phi)\right] &= \sqrt{T}\left[\left[\widehat{N}_T\left(\widehat{\phi}\right)-\widehat{N}_T(\phi)\right]+\left[\widehat{N}_T(\phi)-\phi\right]\right]\end{aligned} \tag{3.88}$$

Define $\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right) = \int_0^1 \widehat{V}_T\left(\widehat{\phi}+\xi\left(\widehat{\phi}-\phi\right)\right) d\xi$, such that the first term on the right-hand side of (3.88) is given by

$$\left[\widehat{N}_T\left(\widehat{\phi}\right)-\widehat{N}_T(\phi)\right] = \widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)\left[\widehat{\phi}-\phi\right] \tag{3.89}$$

Plugging (3.89) into (3.88), the latter reduces to:

$$\begin{aligned}\sqrt{T}\left[\widehat{\phi}-\phi\right] =&\sqrt{T}\left[\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)\left[\widehat{\phi}-\phi\right]\right]+\sqrt{T}\left[\widehat{N}_T\left(\phi\right)-\phi\right]\\ \sqrt{T}\left[\widehat{\phi}-\phi\right] =&\sqrt{T}\left[\left[I_3-\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)\right]^{-1}\left[\widehat{N}_T\left(\phi\right)-\phi\right]\right]\end{aligned} \tag{3.90}$$

As in Dominitz and Sherman (2005), we initially show that $\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right) \xrightarrow{p} V\left(\phi\right)$. To this purpose, write $\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)$ as

$$\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)=V\left(\phi\right)+\left[\Lambda\left(\widehat{\phi},\phi\right)-V\left(\phi\right)\right]+\left[\widehat{\Lambda}_T\left(\widehat{\phi},\phi\right)-\Lambda\left(\widehat{\phi},\phi\right)\right] \tag{3.91}$$

Item i in Theorem 2 states that $\widehat{\phi}$ converges in probability to ϕ as $j \longrightarrow \infty$ with $T \longrightarrow \infty$. This implies that $\Lambda\left(\widehat{\phi},\phi\right) \xrightarrow{p} V\left(\phi\right)$, yielding that the second term on the right-hand of (3.91) converges in probability to zero. Lemma 10 implies that the third term on the right-hand side of (3.91) has order $o_p\left(1\right)$. Hence, (3.90) reduces to

$$\sqrt{T}\left[\widehat{\phi}-\phi\right] =\sqrt{T}\left[\left[I_3-V\left(\phi\right)\right]^{-1}\left[\widehat{N}_T\left(\phi\right)-\phi\right]\right] \tag{3.92}$$

It remains to study the asymptotic distribution of $\sqrt{T}\left[\widehat{N}_T\left(\phi\right)-\phi\right]$. Note that, when $T \longrightarrow \infty$ and $N_T\left(.\right)$ is evaluated on the true vector of parameter, the sample mapping reduces, asymptotically, to the case where the latent variable becomes observed regressors. Given that, the asymptotic distribution of $\sqrt{T}\left[\widehat{N}_T\left(\phi\right)-\phi\right]$ reduces to the asymptotic distribution of the NL-LS estimator. As in Greene (2008), the asymptotic variance of the NL-LS estimator is given by $\sigma_u^2 C_0^{-1}$, where $C_0 = \text{plim}\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial h_t(\theta)}{\partial\theta}\frac{\partial h_t(\theta)}{\partial\theta'}\right]$ and $h_t\left(\theta\right)$ is the nonlinear function in $Q_T\left(y_t,x_t;\theta\right)=\left(y_t-h_t\left(x_t,\theta\right)\right)^2$. Considering the sample mapping of the NL-ILS estimator, the function $h_t\left(x_t,\phi\right)$

is given by the MA($\bar{q}$), such that $h_t(u_t, \phi) = \psi_0 + \sum_{i=0}^{\bar{q}} \psi_i u_{t-1-i}$. Given that, C_0 resumes to

$$C_0 = \begin{pmatrix} \frac{1}{(1-a)^2} & -\frac{\omega}{(1-a)^3} & 0 \\ -\frac{\omega}{(1-a)^3} & \frac{\omega^2}{(1-a)^4} + \sum_{i=0}^{\bar{q}} d_i^2 \sigma_u^2 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 \\ 0 & \sum_{i=0}^{\bar{q}} d_i a^i \sigma_u^2 & \sum_{i=0}^{\bar{q}} a^{2i} \sigma_u^2 \end{pmatrix} \tag{3.93}$$

where $d_i = ia^{i-1}(a+b) + a^i$. Applying the central limit theorem for martingale difference sequences, the asymptotic distribution of $\sqrt{T}\left[\widehat{N}_T(\phi) - \phi\right]$ is given by

$$\sqrt{T}\left[\widehat{N}_T(\phi) - \phi\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2 C_0^{-1}\right) \tag{3.94}$$

Equation 3.67 gives the analytical solution of $V(\phi)$. Define $A = [I - V(\phi)]^{-1}$, then the asymptotic distribution of the NL-ILS is given by

$$\sqrt{T}\left[\widehat{\phi} - \phi\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2 A C_0^{-1} A'\right) \tag{3.95}$$

■

Proof of Corollary 1: This proof follows the item (i) in Theorem 2. Note that Lemma 7 holds because u_t has zero mean, finite variance and autocovariance equal to zero for all lags greater than zero. It is relevant to discuss the validity of Lemmas 10 and 9. Both of them are based on the weak law of large numbers. Note that, from item (i) in corollary 1, the u_t is a linear projection with $\text{Cov}(u_{t-i}, u_{t-j}) = 0$ for all $i \neq j$. This is sufficient to allow

the use of the weak law of large numbers as stated in Hamilton (1994) - pg. 186, implying that Lemma 9 holds. If Lemma 9 holds, then Lemma 10 also holds, extending the validity of item (i) in Theorem 2 to this Corollary.

Proof of Proposition 1: The regressors in (3.26) do not depend on $\widehat{\lambda}_{j+1}$. This implies that the first derivative of the sample mapping with respect to $\widehat{\lambda}_{j+1}$ is

$$4\frac{1}{T}\sum_{t=1}^{T}\left\{\left[\left[y_t-\widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right]-\widehat{\psi}_{j+1,0}-\sum_{i=0}^{\bar{q}}\widehat{\psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2\times\right.$$
$$\left.\left[y_t-\widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right]\widehat{\sigma}_{j,t}\right\}=0$$
$$\frac{1}{T}\sum_{t=1}^{T}\left\{\left[y_t-\widehat{\lambda}_{j+1}\widehat{\sigma}_{j,t}\right]\widehat{\sigma}_{j,t}\right\}=0 \tag{3.96}$$

By manipulating (3.96), $\widehat{\lambda}_{j+1}$ resumes to:

$$\widehat{\lambda}_{j+1}=\left[\sum_{t=1}^{T}\widehat{\sigma}_{j,t}^2\right]^{-1}\sum_{t=1}^{T}\widehat{\sigma}_{j,t}y_t \tag{3.97}$$

The remaining first order conditions do not have a closed solution, implying that $\widehat{\phi}_{j+1}$ has to be recovers through optimization. This concludes the proof of Proposition 1.

Figure 3.1: GARCH(1,1): ACM property

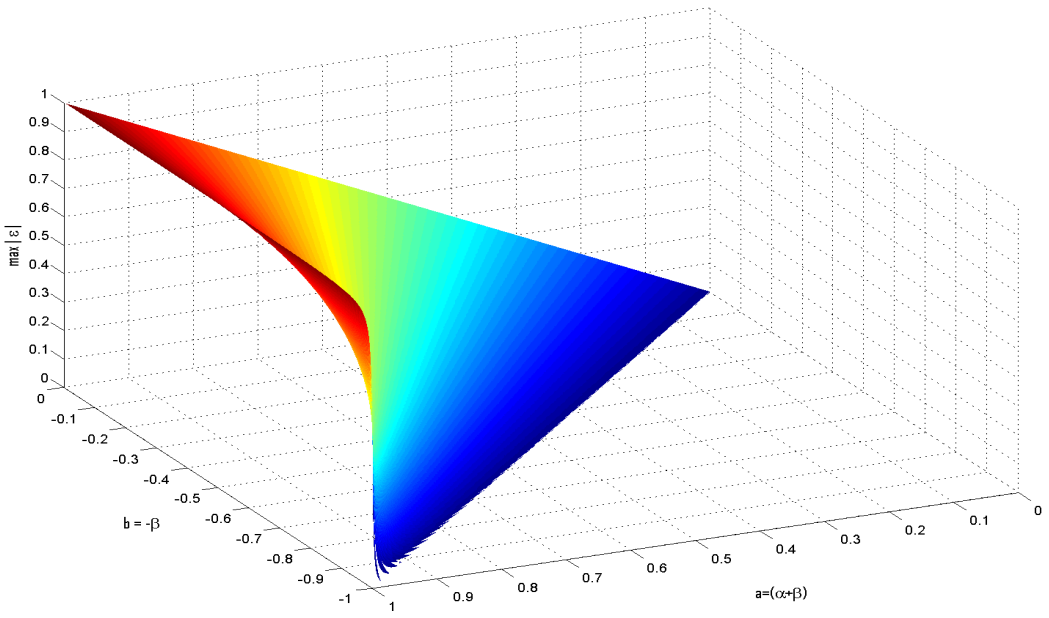


Figure 3.1 plots the highest element of $|\varepsilon|$ in (3.68) using different combinations of α and β, such that Assumption B1 is satisfied. The grid is fixed in 0.001.

Figure 3.2: GARCH(1,1)-in-mean: ACM property

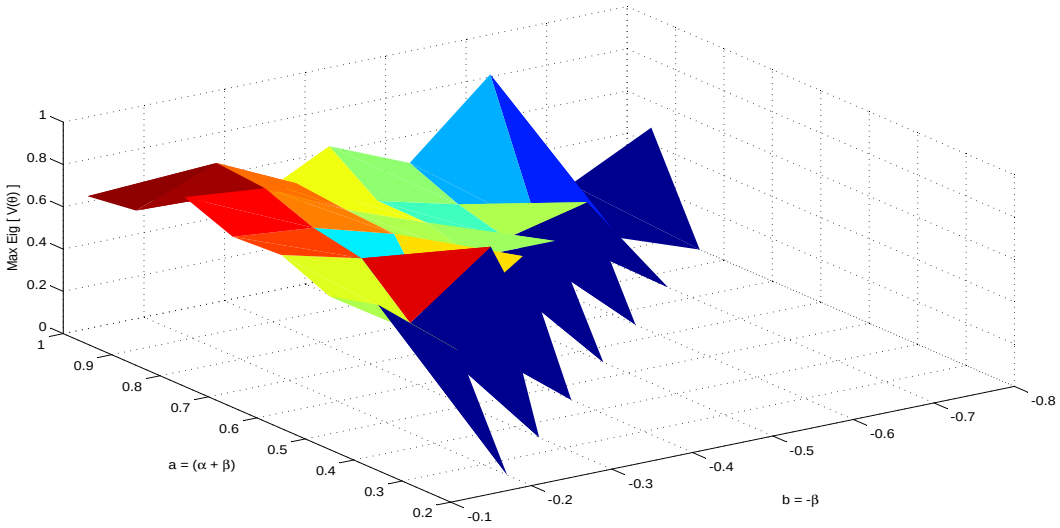


Figure 3.2 displays the maximum eigenvalue computed from the numerical gradient of the NL-ILS mapping.

Table 3.1: GARCH(1,1)

		T = 100			T = 200			T = 300			T = 500		
		NL-ILS	MLE		NL-ILS	MLE		NL-ILS	MLE		NL-ILS	MLE	
		Median	Median	RRMSE	Median	Median	RRMSE	Median	Median	RRMSE	Median	Median	RRMSE
ω	0.010	0.002	0.058	0.391	0.002	0.048	0.461	0.003	0.039	0.599	0.003	0.028	0.773
α	0.025	0.035	0.050	0.562	0.026	0.033	0.684	0.023	0.028	0.760	0.021	0.026	0.910
β	0.970	0.961	0.908	0.390	0.971	0.939	0.535	0.974	0.950	0.705	0.976	0.958	0.925

		T = 100			T = 200			T = 300			T = 500		
		NL-ILS	MLE		NL-ILS	MLE		NL-ILS	MLE		NL-ILS	MLE	
		Median	Median	RRMSE	Median	Median	RRMSE	Median	Median	RRMSE	Median	Median	RRMSE
ω	0.010	0.001	0.033	0.911	0.002	0.021	1.128	0.005	0.017	1.258	0.006	0.014	1.308
α	0.025	0.038	0.067	0.543	0.031	0.042	0.737	0.028	0.034	0.855	0.026	0.030	0.977
β	0.945	0.941	0.836	0.788	0.952	0.892	1.002	0.954	0.915	1.161	0.956	0.926	1.226

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 5000 replications. Replications that do not achieve convergence are discarded when computing the relative measures.

Table 3.2: weak-GARCH(1,1): Specification 1

	m=2, T = 500					m=3, T = 333			
			NL-ILS	MLE			NL-ILS	MLE	
	High Freq.	Low freq.	Median	Median	RRMSE	Low freq.	Median	Median	RRMSE
ω	0.001	0.002	0.002	0.003	1.030	0.003	0.003	0.005	1.000
α	0.050	0.065	0.048	0.065	1.277	0.074	0.056	0.078	1.050
β	0.940	0.915	0.929	0.900	1.081	0.896	0.908	0.868	0.970

	m=4, T = 250					m=5, T = 200			
			NL-ILS	MLE			NL-ILS	MLE	
	High Freq.	Low freq.	Median	Median	RRMSE	Low freq.	Median	Median	RRMSE
ω	0.001	0.004	0.005	0.007	0.904	0.005	0.005	0.008	0.898
α	0.050	0.081	0.059	0.086	0.962	0.086	0.060	0.093	0.827
β	0.940	0.879	0.891	0.835	0.882	0.865	0.880	0.809	0.854

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of the median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded when computing the relative measures. The variable m denotes sampling frequency.

Table 3.3: weak-GARCH(1,1): Specification 2

	m=2, T = 500					m=3, T = 333			
			NL-ILS	MLE			NL-ILS	MLE	
	High Freq.	Low freq.	Median	Median	RRMSE	Low freq.	Median	Median	RRMSE
ω	0.001	0.002	0.001	0.004	0.887	0.003	0.001	0.006	0.826
α	0.020	0.024	0.026	0.032	0.885	0.027	0.027	0.039	0.789
β	0.970	0.956	0.956	0.927	0.868	0.943	0.960	0.901	0.805

	m=4, T = 250					m=5, T = 200			
			NL-ILS	MLE			NL-ILS	MLE	
	High Freq.	Low freq.	Median	Median	RRMSE	Low freq.	Median	Median	RRMSE
ω	0.001	0.004	0.002	0.009	0.848	0.005	0.001	0.011	0.807
α	0.020	0.028	0.033	0.046	0.719	0.029	0.035	0.057	0.709
β	0.970	0.932	0.943	0.850	0.815	0.922	0.945	0.824	0.792

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. Results reported in terms of the median and the relative root mean squared error (RRMSE) for the conditional variance parameters. RRMSE is computed using the RMSE from the QMLE estimates on the denominator and the RMSE of the NL-ILS estimates on the numerator. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded when computing the relative measures. The variable m denotes sampling frequency.

Table 3.4: GARCH(1,1)-in-mean

	T = 100		T = 200		T = 300		T = 400		T = 500		T = 750	
	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.1$	1.25	0.96	1.11	1.00	1.11	0.99	1.13	1.00	1.09	1.01	1.03	1.00
$\omega = 0.01$	0.16	2.54	0.18	1.23	0.23	1.12	0.36	1.05	0.41	1.04	0.68	1.01
$\alpha = 0.025$	0.85	1.48	0.68	0.99	0.87	0.97	1.05	0.96	1.16	0.96	1.36	0.99
$\beta = 0.97$	0.60	1.83	0.63	1.28	0.70	1.13	1.00	1.05	1.12	1.02	1.57	1.01
min $\widehat{\pi}_{t+h}$	1.23	1.00	1.09	1.00	1.09	0.99	1.13	1.00	1.12	1.00	1.10	1.00
med $\widehat{\pi}_{t+h}$	1.25	1.01	1.10	1.00	1.10	0.99	1.14	1.00	1.13	1.00	1.12	1.00
max $\widehat{\pi}_{t+h}$	1.25	1.03	1.11	1.00	1.11	1.00	1.15	1.00	1.14	1.01	1.12	1.00
min σ^2_{t+h}	0.48	0.70	0.72	0.95	0.87	0.97	1.02	0.97	1.07	0.99	1.11	0.99
med σ^2_{t+h}	0.52	0.74	0.78	0.97	0.90	0.98	1.02	0.98	1.08	0.99	1.11	1.00
max σ^2_{t+h}	0.57	0.76	0.81	0.98	0.93	0.99	1.04	0.99	1.10	1.00	1.14	1.00

	T = 100		T = 200		T = 300		T = 400		T = 500		T = 750	
	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.1$	1.20	1.01	1.11	0.99	1.11	1.04	1.13	0.99	1.03	1.01	1.06	1.02
$\omega = 0.01$	2.28	2.67	1.20	1.43	1.46	1.00	1.22	1.11	1.41	1.12	1.76	1.08
$\alpha = 0.025$	0.93	1.70	0.66	1.00	0.83	0.99	0.87	0.97	0.97	0.96	1.00	0.98
$\beta = 0.945$	1.65	2.39	0.99	1.10	1.22	1.07	1.41	1.09	1.52	1.11	1.70	1.06
min $\widehat{\pi}_{t+h}$	1.31	0.99	1.19	1.03	1.12	1.03	1.06	0.99	1.04	1.02	1.04	1.01
med $\widehat{\pi}_{t+h}$	1.31	1.00	1.20	1.04	1.14	1.05	1.07	0.99	1.06	1.02	1.05	1.01
max $\widehat{\pi}_{t+h}$	1.32	1.00	1.20	1.06	1.14	1.06	1.08	1.00	1.08	1.03	1.06	1.02
min σ^2_{t+h}	0.59	0.72	0.84	0.90	0.95	0.96	1.03	0.99	1.07	0.98	1.10	0.99
med σ^2_{t+h}	0.63	0.76	0.87	0.91	0.99	0.98	1.04	1.00	1.09	0.99	1.11	1.00
max σ^2_{t+h}	0.67	0.81	0.91	0.94	1.03	0.99	1.05	1.00	1.15	1.00	1.16	1.00

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the GARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the relative root median squared error (RMedSE). Forecast accuracy is accessed through the RRMedSFE (relative root median squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.

Table 3.5: GARCH(1,1)-in-mean

	T = 100		T = 200		T = 300		T = 400		T = 500		T = 750	
	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.1$	1.14	0.95	1.06	1.00	1.10	1.02	1.20	1.01	1.09	1.01	1.19	1.00
$\omega = 0.01$	0.34	1.12	0.75	1.12	1.21	1.11	1.47	1.07	1.79	1.08	2.33	1.02
$\alpha = 0.08$	0.88	0.93	1.22	1.03	1.39	1.03	1.56	1.02	1.64	1.00	1.92	1.00
$\beta = 0.90$	0.80	1.01	1.37	1.07	1.77	1.11	2.14	1.09	2.16	1.04	2.78	1.05
min $\widehat{\pi}_{t+h}$	1.18	1.00	1.09	1.00	1.12	1.01	1.15	1.01	1.11	0.99	1.16	0.99
med $\widehat{\pi}_{t+h}$	1.19	1.01	1.11	1.01	1.13	1.02	1.18	1.01	1.13	1.00	1.19	1.00
max $\widehat{\pi}_{t+h}$	1.21	1.02	1.12	1.01	1.15	1.03	1.20	1.02	1.15	1.00	1.22	1.00
min σ^2_{t+h}	0.93	1.04	1.02	1.02	1.05	1.01	1.03	1.01	1.08	1.00	1.08	1.00
med σ^2_{t+h}	0.95	1.05	1.03	1.03	1.07	1.02	1.08	1.01	1.10	1.01	1.12	1.01
max σ^2_{t+h}	0.98	1.07	1.06	1.05	1.09	1.03	1.10	1.02	1.16	1.01	1.16	1.01

	T = 100		T = 200		T = 300		T = 400		T = 500		T = 750	
	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.1$	1.10	1.02	1.15	1.00	1.10	1.00	1.11	1.00	1.08	1.00	1.08	1.00
$\omega = 0.01$	0.89	1.31	1.53	1.16	1.59	1.19	1.86	1.12	2.20	1.10	2.20	1.10
$\alpha = 0.08$	0.94	1.00	1.09	1.06	1.12	1.03	1.27	1.03	1.30	1.05	1.30	1.05
$\beta = 0.875$	0.98	1.12	1.51	1.16	1.67	1.20	1.82	1.13	1.98	1.10	1.98	1.10
min $\widehat{\pi}_{t+h}$	1.10	1.03	1.11	1.02	1.07	0.99	1.04	0.98	1.06	1.01	1.06	1.01
med $\widehat{\pi}_{t+h}$	1.11	1.04	1.12	1.03	1.07	1.00	1.06	0.99	1.09	1.01	1.09	1.01
max $\widehat{\pi}_{t+h}$	1.12	1.05	1.13	1.03	1.08	1.00	1.08	0.99	1.11	1.02	1.11	1.02
min σ^2_{t+h}	0.95	1.01	1.04	1.02	1.03	1.00	1.02	0.97	1.03	1.00	1.03	1.00
med σ^2_{t+h}	0.98	1.03	1.05	1.02	1.04	1.01	1.03	0.99	1.05	1.00	1.05	1.00
max σ^2_{t+h}	1.01	1.05	1.07	1.05	1.05	1.02	1.05	0.99	1.08	1.01	1.08	1.01

NL-ILS and MLE account for results obtained using the NL-ILS and MLE algorithms. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the GARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the relative root median squared error (RMedSE). Forecast accuracy is accessed through the RRMedSFE (relative root median squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.

Table 3.6: RealGARCH(1,1)-in-mean

	T = 100	T = 200	T = 300	T = 400	T = 500	T = 2000
	NL-ILS	NL-ILS	NL-ILS	NL-ILS	NL-ILS	NL-ILS
$\lambda = 0.1$	1.21 (0.099)	1.41 (0.105)	1.63 (0.104)	1.71 (0.103)	1.81 (0.104)	2.35 (0.102)
$\omega = 0.06$	1.74 (0.06)	1.89 (0.05)	1.95 (0.049)	2.00 (0.05)	2.01 (0.053)	2.27 (0.058)
$\beta = 0.45$	1.65 (0.393)	2.45 (0.423)	2.83 (0.442)	3.30 (0.442)	3.70 (0.444)	5.68 (0.458)
$\gamma = 0.51$	1.80 (0.519)	2.63 (0.506)	3.10 (0.496)	3.49 (0.501)	3.84 (0.505)	5.39 (0.501)
$\xi = -0.18$	1.78 (−0.318)	2.03 (−0.227)	2.10 (−0.22)	2.19 (−0.207)	2.21 (−0.209)	2.08 (−0.189)
$\varphi = 1.04$	2.01 (0.942)	2.40 (1.012)	2.54 (1.017)	2.60 (1.021)	2.65 (1.027)	2.51 (1.038)
$\tau_1 = -0.11$	1.03 (−0.119)	1.03 (−0.118)	1.07 (−0.121)	1.06 (−0.122)	1.10 (−0.121)	1.37 (−0.122)
$\tau_2 = 0.07$	1.03 (0.073)	1.11 (0.073)	1.16 (0.076)	1.14 (0.074)	1.21 (0.076)	1.14 (0.073)
$\widehat{\sigma}_t^2$	1.00	0.94	0.96	0.95	0.96	1.00
min $\widehat{\pi}_{t+h}$	1.10	1.20	1.24	1.27	1.28	1.27
med $\widehat{\pi}_{t+h}$	1.14	1.26	1.34	1.34	1.36	1.33
max $\widehat{\pi}_{t+h}$	1.19	1.35	1.45	1.45	1.50	1.45
min $\widehat{\sigma}_{t+h}^2$	0.98	1.01	1.00	1.01	1.01	0.99
med $\widehat{\sigma}_{t+h}^2$	1.02	1.04	1.01	1.02	1.03	1.00
max $\widehat{\sigma}_{t+h}^2$	1.09	1.07	1.03	1.05	1.05	1.01
min $\widehat{\nu}_{t+h}$	0.92	0.97	0.96	0.99	0.98	0.98
med $\widehat{\nu}_{t+h}$	0.97	0.99	0.97	1.00	1.00	0.99
max $\widehat{\nu}_{t+h}$	1.00	1.01	0.99	1.02	1.01	1.00

Table 3.6 reports the results obtained using the NL-ILS a estimator. Results for the RealGARCH(1,1)-in-mean parameters and in-sample conditional variance are reported in terms of the Relative root Mean Squared Error (RMSE). Values inside the brackets refer to the median computed within all valid replications. Forecast accuracy is accessed through the RRMedSFE (relative root median squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.

Table 3.7: Robustness analysis: conditional variance misspecification

	T = 500		T = 750		T = 1000		T = 1250		T = 1500		T = 1750	
APARCH(1,1)-in-mean	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.20$	0.98 (−0.011)	1.00 (−0.032)	0.93 (−0.009)	1.02 (−0.034)	0.88 (−0.008)	1.00 (−0.033)	0.83 (−0.007)	0.97 (−0.035)	0.81 (−0.008)	1.00 (−0.035)	0.78 (−0.007)	1.01 (−0.035)
min $\widehat{\pi}_{t+h}$	1.12	0.99	1.13	0.98	1.07	1.01	1.00	0.99	0.96	1.00	0.93	0.99
med $\widehat{\pi}_{t+h}$	1.16	1.00	1.19	0.98	1.16	1.01	1.11	0.99	1.07	1.00	1.05	0.99
max $\widehat{\pi}_{t+h}$	1.24	1.00	1.31	0.99	1.28	1.02	1.24	0.99	1.21	1.00	1.19	0.99
min $\widehat{\sigma}^2_{t+h}$	1.29	0.96	1.32	0.93	1.35	0.98	1.36	1.01	1.21	0.98	1.21	0.95
med $\widehat{\sigma}^2_{t+h}$	1.52	0.96	1.61	0.94	1.65	1.00	1.64	1.03	1.42	0.99	1.41	0.96
max $\widehat{\sigma}^2_{t+h}$	1.73	0.97	1.90	0.95	1.96	1.03	1.93	1.03	1.68	1.00	1.66	0.97

	T = 500		T = 750		T = 1000		T = 1250		T = 1500		T = 1750	
EGARCH(1,1,1)-in-mean	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.20$	0.75 (0.019)	1.00 (0.049)	0.70 (0.012)	1.00 (0.048)	0.63 (0.009)	0.99 (0.048)	0.61 (0.006)	1.00 (0.047)	0.57 (0.004)	1.00 (0.046)	0.53 (0.003)	1.00 (0.035)
min $\widehat{\pi}_{t+h}$	0.88	1.00	0.83	1.01	0.76	0.98	0.73	0.99	0.71	0.99	0.70	0.99
med $\widehat{\pi}_{t+h}$	0.88	1.01	0.83	1.01	0.78	0.98	0.74	0.99	0.73	0.99	0.72	1.00
max $\widehat{\pi}_{t+h}$	0.89	1.01	0.84	1.01	0.80	0.99	0.76	1.00	0.76	1.00	0.75	1.00
min $\widehat{\sigma}^2_{t+h}$	1.18	1.00	1.11	1.00	1.11	0.99	1.09	0.96	1.07	0.98	1.06	0.97
med $\widehat{\sigma}^2_{t+h}$	1.27	1.01	1.20	1.00	1.20	0.99	1.19	0.97	1.14	0.99	1.14	0.98
max $\widehat{\sigma}^2_{t+h}$	1.33	1.01	1.26	1.01	1.26	0.99	1.25	0.98	1.19	0.99	1.21	0.99

NL-ILS accounts for results obtained using the NL-ILS algorithm. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the λ parameter are reported in terms of the Relative Mean Squared Error (RMSE). Values inside brackets refer to the bias obtained for the λ estimates. Forecast accuracy is accessed through the RRMSFE (relative root mean squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.

Table 3.8: Robustness analysis: conditional variance misspecification

	T = 500		T = 750		T = 1000		T = 1250		T = 1500		T = 1750	
GJR GARCH(1,1,1)-in-mean	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.20$	1.02 (0.004)	1.00 (0.017)	1.01 (0.002)	1.00 (0.016)	1.00 (0)	1.00 (0.014)	0.95 (0.002)	1.00 (0.015)	0.96 (−0.001)	1.00 (0.014)	0.93 (0.001)	1.00 (0.015)
min $\widehat{\pi}_{t+h}$	0.98	1.00	0.96	1.00	0.93	1.00	0.91	1.00	0.90	1.00	0.87	1.00
med $\widehat{\pi}_{t+h}$	1.02	1.00	1.01	1.00	1.01	1.00	0.98	1.00	0.99	1.00	0.95	1.00
max $\widehat{\pi}_{t+h}$	1.05	1.00	1.04	1.00	1.05	1.00	1.01	1.00	1.02	1.00	0.98	1.00
min $\widehat{\sigma}^2_{t+h}$	1.06	0.99	1.06	1.00	1.06	1.00	1.04	1.00	1.05	1.00	1.04	1.00
med $\widehat{\sigma}^2_{t+h}$	1.12	1.00	1.13	1.00	1.13	1.00	1.09	1.00	1.10	1.00	1.08	1.00
max $\widehat{\sigma}^2_{t+h}$	1.28	1.00	1.30	1.00	1.30	1.00	1.22	1.00	1.28	1.00	1.22	1.00

	T = 500		T = 750		T = 1000		T = 1250		T = 1500		T = 1750	
GARCH(2,2)-in-mean	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*	NL-ILS	MLE*
$\lambda = 0.20$	1.13 (0.003)	1.00 (0.001)	1.13 (0.003)	1.00 (0)	1.15 (0.001)	1.01 (0)	1.13 (0.002)	1.00 (0.001)	1.12 (0.003)	0.99 (0.029)	1.11 (0.002)	1.00 (0.001)
min $\widehat{\pi}_{t+h}$	1.16	1.00	1.14	1.00	1.17	1.01	1.15	0.99	1.14	1.00	1.13	1.00
med $\widehat{\pi}_{t+h}$	1.17	1.00	1.16	1.00	1.20	1.01	1.19	1.00	1.17	1.00	1.16	1.00
max $\widehat{\pi}_{t+h}$	1.20	1.00	1.19	1.00	1.25	1.01	1.23	1.00	1.25	1.00	1.23	1.00
min $\widehat{\sigma}^2_{t+h}$	1.17	1.00	1.13	0.99	1.15	0.99	1.16	0.99	1.14	0.99	1.14	0.99
med $\widehat{\sigma}^2_{t+h}$	1.27	1.01	1.25	1.00	1.27	1.00	1.30	1.00	1.28	1.00	1.28	1.00
max $\widehat{\sigma}^2_{t+h}$	1.48	1.01	1.52	1.01	1.68	1.00	1.81	1.00	1.82	1.00	1.87	1.00

NL-ILS accounts for results obtained using the NL-ILS algorithm. MLE* accounts for results obtained using the MLE estimator computed using NL-ILS estimates as the initial values. Results for the λ parameter are reported in terms of the relative root mean squared error (RMSE). Values inside brackets refer to the bias obtained for the λ estimates. Forecast accuracy is accessed through the RRMSFE (relative root mean squared forecast error). Relative measures are computed with respect to the MLE benchmark. Relative measures less than one imply NL-ILS estimator outperforms the MLE methodology. Truncation parameter is fixed to $\bar{q} = 3\sqrt[4]{T}$. We perform 1500 replications. Replications that do not achieve convergence are discarded for computing the relative measures.

Table 3.9: Descriptive statistics

	Mean	Median	Std. Dev.	Kurtosis	N. Obs	Start Date	End Date
CRSP†	0.0002	0.0005	0.0099	19.7	12,148	28/06/1963	29/09/2011
S&P500†	0.0000	0.0002	0.0104	31.1	12,148	28/06/1963	29/09/2011
S&P100†	0.0001	0.0004	0.0121	30.5	7,364	04/08/1982	29/09/2011
CRSP‡	0.0010	0.0026	0.0227	9.0	2,426	05/07/1963	30/09/2011
S&P500‡	0.0001	0.0010	0.0227	11.6	2,426	05/07/1963	30/09/2011
S&P100‡	0.0007	0.0019	0.0243	8.3	1,469	04/08/1982	29/09/2011
CRSP§	0.0061	0.0096	0.0545	10.4	1,023	01/07/1926	01/08/2011
S&P500§	0.0019	0.0054	0.0424	5.3	740	01/01/1950	01/08/2011
S&P100§	0.0023	0.0061	0.0492	7.1	330	02/04/1984	02/10/2011

Superscripts †, ‡ and § denote daily, weekly and monthly frequencies, respectively. The null hypothesis in the Jarque-Bera test is reject in all indices and frequencies.

Table 3.10: Empirical application: risk premium estimation

	Daily freq.					
	CRSP		S&P500		S&P100	
	NL-ILS	QMLE	NL-ILS	QMLE	NL-ILS	QMLE
λ	0.02^{*} (0.013)	0.07^{***} (0.009)	0.01 (0.010)	0.04^{***} (0.009)	0.02^{*} (0.012)	0.05^{***} (0.011)
ω	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.000)	0.00^{***} (0.000)
α	0.12^{***} (0.023)	0.09^{***} (0.002)	0.10^{***} (0.022)	0.08^{***} (0.002)	0.11^{***} (0.015)	0.08^{***} (0.002)
β	0.84^{***} (0.024)	0.91^{***} (0.003)	0.83^{***} (0.034)	0.92^{***} (0.002)	0.80^{***} (0.077)	0.91^{***} (0.003)

	Weekly freq.					
	CRSP		S&P500		S&P100	
	NL-ILS	QMLE	NL-ILS	QMLE	NL-ILS	QMLE
λ	0.06^{**} (0.024)	0.11^{***} (0.019)	0.01 (0.022)	0.06^{***} (0.020)	0.03 (0.029)	0.08^{***} (0.025)
ω	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.000)	0.00^{***} (0.000)
α	0.12^{***} (0.042)	0.14^{***} (0.011)	0.11^{**} (0.058)	0.13^{***} (0.009)	0.14^{***} (0.054)	0.14^{***} (0.012)
β	0.77^{***} (0.209)	0.84^{***} (0.013)	0.81^{***} (0.204)	0.85^{***} (0.012)	0.74^{***} (0.207)	0.84^{***} (0.015)

	Monthly freq.					
	CRSP		S&P500		S&P100	
	NL-ILS	QMLE	NL-ILS	QMLE	NL-ILS	QMLE
λ	0.12^{***} (0.045)	0.18^{***} (0.031)	0.05 (0.044)	0.07^{*} (0.038)	0.05 (0.063)	0.06 (0.058)
ω	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.000)	0.00^{***} (0.000)	0.00 (0.001)	0.00^{*} (0.000)
α	0.09 (0.063)	0.14^{***} (0.019)	0.11^{***} (0.042)	0.11^{***} (0.025)	0.04 (0.078)	0.14^{***} (0.042)
β	0.88^{***} (0.138)	0.84^{***} (0.018)	0.71^{***} (0.186)	0.85^{***} (0.028)	0.77^{***} (0.218)	0.82^{***} (0.055)

Standard errors are reported inside the brackets. NL-ILS standard errors are obtained using block bootstrap algorithm with 1000 replications. QMLE standard errors are computed using Bollerslev-Wooldridge robust estimator. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.

Table 3.11: Empirical application: risk premium estimation - RealGARCH(1,1)-in-mean

	S&P500					
	Daily freq.		Weekly freq.		Monthly freq.	
	NL-ILS	MLE	NL-ILS	MLE	NL-ILS	MLE
λ	-0.03 (0.019)	0.01 (0.020)	-0.03 (0.052)	-0.02 (0.045)	-0.10 (0.115)	-0.05 (0.110)
ω	1.44^{***} (0.242)	0.59^{***} (0.165)	0.50 (0.493)	-0.23 (0.521)	-0.66 (0.941)	-0.75 (1.277)
β	0.51^{***} (0.045)	0.56^{***} (0.032)	0.29^{***} (0.074)	0.28^{***} (0.053)	0.25^{**} (0.109)	0.28^{**} (0.130)
γ	0.63^{***} (0.061)	0.49^{***} (0.041)	0.75^{***} (0.086)	0.65^{***} (0.071)	0.62^{***} (0.150)	0.57^{***} (0.152)
ξ	-2.61^{***} (0.284)	-1.84^{***} (0.306)	-1.26^{*} (0.645)	-0.61 (0.748)	-0.11 (2.164)	-0.16 (2.360)
φ	0.75^{***} (0.030)	0.83^{***} (0.034)	0.88^{***} (0.083)	0.98^{***} (0.104)	1.03^{***} (0.348)	1.02^{**} (0.401)
τ_1	-0.14^{***} (0.012)	-0.15^{***} (0.012)	-0.20^{***} (0.026)	-0.22^{***} (0.032)	-0.31^{***} (0.062)	-0.32^{***} (0.073)
τ_2	0.00 (0.010)	0.01 (0.012)	0.07^{***} (0.022)	0.08^{***} (0.022)	0.08^{*} (0.045)	0.08^{**} (0.036)

Standard errors are reported inside the brackets. NL-ILS and MLE standard errors are obtained using block bootstrap algorithm with 1000 replications. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.

Table 3.12: Robustness check: risk premium estimation

	EGARCH(1,1,1)-in-mean			APARCH(1,1,1)-in-mean			GJR-GARCH(1,1,1)-in-mean		
	CRSP	S&P500	S&P100*	CRSP	S&P500	S&P100*	CRSP	S&P500	S&P100*
Daily Freq.	0.031^{***}	0.004	0.013	0.033^{***}	0.006	0.015	0.039^{***}	0.012	0.022^{*}
	(0.009)	(0.009)	(0.012)	(0.009)	(0.009)	(0.012)	(0.009)	(0.009)	(0.012)
Weekly Freq.	0.064^{***}	0.004	0.041	0.062^{***}	0.005	0.042	0.069^{***}	0.017	0.042
	(0.021)	(0.020)	(0.026)	(0.021)	(0.021)	(0.026)	(0.021)	(0.021)	(0.026)
Monthly Freq.	0.152^{***}	0.065^{*}	0.051	0.152^{***}	0.014	0.015	0.156^{***}	0.063	0.060
	(0.032)	(0.039)	(0.061)	(0.032)	(0.033)	(0.012)	(0.032)	(0.039)	(0.062)

We report estimates of λ computed using QMLE methodology. Standard errors are computed using the Bollerslev-Wooldridge estimator. The symbols *, **, and *** denote significance 10%, 5% and 1%, respectively.

Chapter 4

Inference on GARCH-in-mean models with time-varying coefficients: assessing risk premium over time

4.1 Introduction

Time-varying volatility plays a major role in both finance and economics. In special, asset return volatility is paramount in fields such as asset pricing, risk management and portfolio allocation. The task of modeling the conditional variance has been a central topic in econometrics following the seminal papers of Engle (1982) and Bollerslev (1986). Since then, different specifications and frameworks, such as GARCH-type models, stochastic volatility, realized volatility and combinations of these approaches have been adopted, trying to capture the very specific stylized facts observed

in financial returns. A natural extension that emerges from modeling the conditional variance is the relation between risk and return. The intertemporal capital asset pricing model (ICAPM) establishes a positive relation between the conditional excess returns and the conditional variance, implying that investors should be remunerated for bearing extra risk. To assess the risk-return tradeoff postulated by the ICAPM model, Engle, Lilien, and Robins (1987) formulates the (G)ARCH-in-mean specification, where a function of the latent conditional variance appears in the mean equation as a regressor. Following Engle, Lilien, and Robins (1987)'s work, the risk-return tradeoff literature has rapidly evolved, however empirical evidences on the sign and significance of the risk premium parameter remain blurred. The justification for these mixed empirical evidences lies on three different issues: first, misspecification of the risk premium function; second, misspecification of the conditional variance equation; third, use of only a few conditioning variables.

In this chapter, we undertake inference on the risk-return tradeoff by using an econometric framework that encompasses the three issues previously discussed. We firstly address the misspecification of the risk premium function by modelling the risk premium parameter as a time-varying stochastic process. To this purpose, we introduce the time-varying GARCH-in-mean (TVGARCH-in-mean) model, where the risk premium parameter is allowed to evolve as a bounded random walk process. By using such specification, we obtain a stochastic risk premium function that is no longer a deterministic function of the conditional standard deviation. Secondly, by modelling the risk premium parameter as a bounded random walk process, we allow its disturbance term to summarize information from a wide range of latent

variables. The issue of biased estimates of the risk premium parameters that arises from misspecification of the conditional variance is addressed by using a kernel based version of the robust nonlinear iterative least squares (NL-ILS) estimator. Using the excess returns computed using the CRSP index on weekly and monthly frequencies, we document that the risk premium parameter is indeed time-varying, alternating positive and negative values over time. Regarding results obtained with excess returns sampled on weekly frequency, we show that the time-varying risk premium parameter picks on periods that precedes the financial crises and economic recessions, and turns negative during high volatility times. Considering the monthly frequency, we find smoother estimates of the time-varying risk premium parameter, which contributes to narrower confidence intervals and stronger significance analyses. We report that the time-varying risk premium parameter is statistically different from zero on almost half of the observations.

The methodology we adopt in this chapter originates in the applied macroeconomics literature, where the time-varying coefficient models have addressed issues such as structural changes on macroeconomic variables and in particular the Great Moderation phenomenon. Estimation strategies that use kernel methods showed to be valid alternatives on assessing these models. Robinson (1989) and Orbe, Ferreira, and Rodriguez-Poo (2005)) assume that the time-varying coefficient is a deterministic (smooth) function of time, whereas Giraitis, Kapetanios, and Yates (2010) model it as a bounded random walk process. We construct the kernel based NL-ILS estimator using the theoretical insights developed by the latter authors.

With regard to previous studies in the risk-return literature, we split

these results in three different groups: first, the full parametric GARCH-in-mean class of models, which includes other parametric specifications of the conditional variance such as EGARCH, GJR-GARCH and stochastic volatility models. Second, the semiparametric GARCH-in-mean models, where estimates of the conditional variance are obtained through a parametric specification of the conditional variance, whereas risk premium function is estimated using nonparametric techniques. Third, models that use measures of realized variance and a broader set of conditioning variables. Also, this third class of models are generic enough to allow for a nonlinear risk premium function.

Considering the first group, mixed evidences in both sign and significance of the time-invariant risk premium parameter have been found in the literature. While French, Schwert, and Stambaugh (1987) find a positive value for λ, Glosten, Jagannathan, and Runkle (1993) find an opposite sign, Baillie and DeGennaro (1990) find very little evidence for a statistically significant λ. We find in Chapter 3 that λ is only significant when the CRSP dataset is adopted. To support this result, we argue that significance analyses using quasi-maximum likelihood (QMLE) estimates of the parameters of GARCH-in-mean models are blurred, following a potential bias associated with the parameters in the mean equation. In fact, if the conditional variance is misspecified, QMLE estimates of the risk premium parameter may be biased, as discussed in Bollerslev, Chou, and Kroner (1992). Furthermore, apart from the work of Christensen, Dahl, and Iglesias (2012), asymptotic theory supporting the use of the QMLE estimator on GARCH-in-mean models is not well established as it is in the GARCH family of models, relying, among others, on the assumption

that the disturbances are martingale difference sequence (m.d.s) processes. This assumption, however, fails when sampling frequency changes, yielding a class of models (weak-GARCH models) which possess disturbances that present some degree of dependence. Chapter 3 shows that the nonlinear iterative least squares (NL-ILS) estimator is robust to misspecification of the conditional variance, delivering unbiased estimates of the risk premium parameter under a variety of volatility specifications. Furthermore, we establish the asymptotic theory considering the GARCH(1,1) case, as well the consistency of the NL-ILS estimator when the disturbances are linear projections (the case of the weak-GARCH(1,1) specification). This chapter adopts the NL-ILS estimator as the core estimation procedure, using therefore the desirable properties associated with the NL-ILS to construct inference on the time-varying risk premium parameter.

With respect to the semiparametric GARCH-in-mean literature, Linton and Perron (2003), Christensen, Dahl, and Iglesias (2012) and Conrad and Mammen (2008) find strong evidences that the risk-return tradeoff is nonlinear, corroborating Pagan and Hong (1990) who argued that the linear relationship between the conditional variance and the excess returns only occurs in very particular cases. Furthermore, Veronesi (2000) shows that the risk premium function can virtually take any form, strengthening the choice of these authors of using the nonparametric framework to recover the risk premium function. Although we use kernel functions to estimate the time-varying risk premium parameter, the TVGARCH-in-mean framework departures from the semiparametric GARCH-in-mean approach in two different directions: firstly, we assume a linear relationship between the conditional standard deviation and λ_t, whereas Linton and Perron (2003),

Christensen, Dahl, and Iglesias (2012) and Conrad and Mammen (2008) assume that the risk premium function is an unknown deterministic function of the conditional variance. Secondly, we admit λ_t to evolve stochastically as an independent process, which allows the risk premium function to depend on exogenous latent shocks and be therefore a stochastic function. This flexible feature of our specification is able to address an important point raised by Lettau and Ludvigson (2010) and Rossi and Timmermann (2010): the disagreement in the risk-return tradeoff literature arises from the use of few conditioning variables and misspecification of the risk premium function. By modelling λ_t as a random walk process, we therefore allow the time-varying risk premium parameter to summarize information from conditioning variables driving the real economy.

A third class of models relies on the use of different datasets that include macroeconomic variables. Lettau and Ludvigson (2010) adopts the dynamic factor analysis, Ghysels, Santa-Clara, and Valkanov (2005) the MIDAS approach and Rossi and Timmermann (2010) the regression trees framework.

It is important to stress that by modelling λ_t as an exogenous stochastic process, we address the points raised by Pagan and Hong (1990), Veronesi (2000) and others, regarding the shape of the risk premium function.

This chapter is organized as follows. Section 4.2 introduces the TVGARCH-in-mean model and the kernel based NL-ILS estimator. By focusing on the TVGARCH(1,1)-in-mean specification, we describe the estimation algorithm as well as the bootstrap methodology we implement to compute the confidence intervals associated with the parameters estimates. Section 4.3 covers the numerical illustrations. We start with the Monte

Carlo study showing that the kernel based NL-ILS estimator presents a good finite sample performance on estimating the time-varying risk premium parameter and the parameters of the conditional variance equation. Finally, we investigate the risk-return tradeoff over time using the excess returns computed using the CRSP index. Section 4.4 concludes. The Appendix brings tables and graphs.

4.2 The time-varying GARCH-in-mean specification

In this section we introduce the TVGARCH-in-mean model as a framework to recover the time-varying risk premium parameter denoted as λ_t, $t = 1, 2, .., T$. Giraitis, Kapetanios, and Yates (2010) established the asymptotic theory for the class of autoregressive models driven by a random drifting autoregressive parameter. We extend their work by allowing the regressors to be latent, which is the case of the TVGARCH-in-mean model. We specify the TVGARCH-in-mean allowing λ_t to evolve stochastically as a bounded random walk process. We start our discussion with a generic specification of the TVGARCH-in-mean model that encompasses specifications with exogenous variables in the conditional variance equation.

These specifications are particularly important because they nest models that use measures of realized variances as regressors. As pointed out by Andersen, Bollerslev, Diebold, and Labys (2003), models that combine the conditional variance with realize measures tend outperform the standard GARCH-type models when forecasting the conditional variance. The

intuition for such results arises because these augmented models tend to respond faster to abrupt changes in the underline volatility than the standard GARCH-type models. Among these models, we highlight the HEAVY, GARCH-X and RealGARCH models proposed by Shepard and Sheppard (2010), Engle (2002) and Hansen, Huang, and Shek (2012), respectively. In principle, the generic model in equations (4.1), (4.3) and (4.4) could be changed in order to accommodate the latter two specifications in spirit of the generic model in Chapter 3. As a matter of simplicity, however, we restrict ourselves to the generic TVGARCH-in-mean model as:

$$y_t = \lambda_t \sigma_t + \epsilon_t \tag{4.1}$$

$$\epsilon_t = \sigma_t \eta_t \tag{4.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \tag{4.3}$$

$$\epsilon_t^2 = \Psi_0 + u_t + \sum_{i=1}^{\infty} \Psi_i u_{t-i} \qquad \Psi_i = \varrho_i(\theta_2), \quad i = 0, 1, ..., \infty \tag{4.4}$$

where σ_t is a latent variable (conditional standard deviation); $\epsilon_t^2 = (y_t - \lambda_t \sigma_t)^2$; u_t is a vector of m.d.s. processes, such that $\mathbb{E}(u_t) = 0$ and $\text{Var}(u_t) = \sigma_u$; $\lambda = (\lambda_1, \lambda_2, ..., \lambda_T)'$, θ_2 is a vector of free parameters in (4.3) and $\theta = (\lambda, \theta_2)'$. The parameter λ_t is known as the risk premium parameter (time-varying in our specification) and the risk premium function, μ_t, is defined as $\mu_t = \lambda_t \sigma_t$.

Similarly as in Giraitis, Kapetanios, and Yates (2010), the time-varying coefficient in (4.1), λ_t, evolves as a rescaled random walk process bounded

between some constant c, such that $-c \leq \lambda_t \leq c$.

$$\lambda_t = c \, \frac{a_t}{\max_{0 \leq \kappa \leq t} |a_\kappa|} \tag{4.5}$$

$$a_t = a_{t-1} + \xi_t \tag{4.6}$$

where ξ_t is a zero mean covariance stationary process with finite fourth moment[1]. Given its parametrization, λ_t is a partial sum of past values of ξ_t, which is set to be orthogonal to σ_t and summarizes information contained in different information sets. Note that under the TVGARCH-in-mean specification, μ_t is allowed to be linear on σ_t but it remains stochastic following the nature of λ_t. Therefore, by allowing μ_t to be stochastic, we departure from the semiparametric GARCH-in-mean specification used by Linton and Perron (2003), Christensen, Dahl, and Iglesias (2012) and Conrad and Mammen (2008).

To estimate the parameters in (4.1), (4.3) and (4.4), we adopt the NL-ILS estimator in the spirit of Chapter 3. The NL-ILS estimator is an iterative estimator that consists on updating recursively, on each iteration, σ_t and then using it to compute the time-varying coefficient λ_t and the remaining parameters θ_2. Denote $\mathbb{B}$ as the space where θ is defined. As in Dominitz and Sherman (2005), we define two mappings: population and sample mappings, which are the solution of the optimization of the population, $\mathbb{E}\left(Q_T\left(y_t;\theta\right)\right)$, and sample, $Q_T\left(y_t;\theta\right)$, objective functions, respectively. To define both objective functions, we adopt the type of smoothed sum of squared residuals target function as discussed in Robinson (1989),

[1] In Section 4.3.1, we discuss how the Kernel based NL-ILS estimator performs when the λ_t is no longer a random walk process, but a stationary AR(1) process taking the form of: $\lambda_t = \phi\lambda_{t-1} + \xi_t$, with $|\phi| < 1$.

Orbe, Ferreira, and Rodriguez-Poo (2005), Kapetanios (2008) and Giraitis, Kapetanios, and Yates (2010). The two mappings are therefore given by:

$$\theta_{j+1} = N(\theta_j) = \min_{\theta_{j+1}} \mathbb{E}\left\{ \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\left(y_\kappa - \lambda_{j+1,t}\sigma_{j,\kappa}\right)^2 - \right.\right. \\ \left.\left. - \Psi_{j+1,0} - \sum_{i=1}^{\infty}\Psi_{j+1,i}u_{j,t-1-i}\right]^2\right\} \tag{4.7}$$

$$\widehat{\theta}_{j+1} = \widehat{N}_T(\widehat{\theta}_j) = \min_{\widehat{\theta}_{j+1}} \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\left(y_\kappa - \widehat{\lambda}_{j+1,t}\widehat{\sigma}_{j,\kappa}\right)^2 - \right. \\ \left. -\widehat{\Psi}_{j+1,0} - \sum_{i=1}^{\bar{q}}\widehat{\Psi}_{j+1,i}\widehat{u}_{j,t-1-i}\right]^2 \tag{4.8}$$

where j accounts for the number of iterations which is a function of T, such that as $T \longrightarrow \infty$, $j \longrightarrow \infty$ at some rate satisfying $\frac{\ln(T)}{j} = o(1)$, $K(x) \geq 0$, $x \in \mathbb{R}$ is kernel function with bounded first derivatives and $\int K(x)\,dx = 1$, H is the bandwidth parameter such that $H \longrightarrow \infty$ and $H = o(T)$; $\Psi_{j+1,i}$, $\widehat{\Psi}_{j+1,i}$ are deterministic functions of $\theta_{2,j+1}$ and $\widehat{\theta}_{2,j+1}$, respectively, and $\bar{q}$ is a truncation parameter, such that $\bar{q} \longrightarrow \infty$ at a logarithmic rate of T. Note that both mappings map from $\mathbb{B}$ to itself, yielding that the iterative procedure is stopped when convergence is achieved. As a identification condition, we have that when evaluated at the true vector of parameters θ, the population mapping returns θ, such that $\theta = N(\theta)$.

Considering the standard GARCH-in-mean specification discussed in Chapter 3, Proposition 1 states that both mappings can be split into two distinct processes: parameters in the mean equation are estimated using ordinary least squares (OLS), whereas the parameters in the conditional

variance are retrieved by adopting the nonlinear least squares (NL-LS) estimator. We generalize the result in Proposition 1 in Chapter 3 to encompass the TVGARCH-in-mean specification. As a result of this, the parameters λ_t in (4.1) are no longer estimated using the OLS estimator, but by using kernel based OLS estimators. The parameters governing the conditional variance equation remain being the estimates obtained using the NL-LS estimator. We formalize this result in Proposition 2.

Proposition 2 *Assume the model stated in (4.1), (4.3) and (4.4). Define the vectors of free parameters in (4.8) on the* $j+1$ *iteration as* $\widehat{\lambda}_{j+1} = (\lambda_{1,j+1}, \lambda_{2,j+1}, ..., \lambda_{T,j+1})'$ *and* $\widehat{\phi}_{j+1} = \left(\widehat{\omega}_{j+1}, \widehat{a}_{j+1}, \widehat{b}_{j+1}\right)'$. *The sample mapping in (4.8) can be computed in two different steps, such that:*

$$i.\ \widehat{\lambda}_{j+1,t} = \left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}_{j,\kappa}^{2}\right]^{-1} \sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}_{j,\kappa} y_{\kappa},\ \textit{for t=1, ...,T}$$

$$ii.\ \widehat{\phi}_{j+1} = \min_{\widehat{\phi}_{j+1}} \sum_{\kappa=1}^{T}\left[\left[y_{\kappa} - \widehat{\lambda}_{j+1,\kappa}\widehat{\sigma}_{j,\kappa}\right]^{2} - \widehat{\psi}_{j+1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{j+1,i}\widehat{u}_{j,\kappa-1-i}\right]^{2}$$

Proof of proposition 2 follows a trivial extension of Proposition 1 in Chapter 3.

Convergence of the NL-ILS estimator relies on the existence of a fixed point, which is determined by the contraction property associated with the mapping. As discussed in Kapetanios (2003), Dominitz and Sherman (2005) and in Chapters 2 and 3, convergence will only occur if the population mapping stated in (4.7) is an Asymptotic Contraction Mapping (ACM)[2]. Furthermore, Chapter 3 argues that the parameters of the infinite

[2]Using the definition in Dominitz and Sherman (2005), a collection $\{K_T^{\omega}(.) : T \geq 1, \omega \in \Omega\}$ is an ACM on $(\mathbb{B}, d)$ if $d\left(K_T^{\omega}(x), K_T^{\omega}(y)\right) \leq cd(x,y)$ as $T \longrightarrow \infty$, where $c \in [0,1)$, $(\mathbb{B}, d)$ is a metric space with $x, y \in \mathbb{B}$, $(\Omega, \mathcal{A}, \mathcal{P})$ denoting a probability space and $K_T^{\omega}(.)$ is a function defined on $\mathbb{B}$.

MA representation of the GARCH component is the bit driving the contraction property of the population mapping of GARCH-in-mean models. Chapter 2 establishes a theoretical bound on the parameters of ARMA(1,1) models that satisfy the ACM condition, whereas Chapter 3 provides analogous results for the GARCH(1,1) case. Using their result, we perform Monte Carlo simulations (available upon request) showing that convergence does not occur in finite sample when the these theoretical bounds are violated. Following that, we implement Monte Carlo validation to assess whether the mapping in (4.8) is an ACM. We show that the NL-ILS algorithm converges for the TVGARCH(1,1)-in-mean model, which supports our claim that the population mapping of the generic TVGARCH-in-mean model is indeed an ACM.

In order to have a rigorous asymptotic inference of the NL-ILS estimator for the TVGARCH(1,1)-in-mean model, it is necessary to combine the theory developed in Dominitz and Sherman (2005) and Giraitis, Kapetanios, and Yates (2010). The first authors provide a generic asymptotic theory for iterative estimators that relies on the contraction property of the population mapping. This condition is proved by evaluating the eigenvalue associated with the theoretical gradient of the population mapping evaluated on the true vector of parameters θ. Additionally to the analytical expression for $V(\theta)$, Theorem 4 in Dominitz and Sherman (2005) requires the asymptotic distribution of the sample mapping evaluated on θ, $\sqrt{T}\left(\widehat{N}_T(\theta) - \theta\right) \xrightarrow{d} N(0, \Sigma)$. To derive this asymptotic result, it is necessary to use Theorem 2.3 in Giraitis, Kapetanios, and Yates (2010), where they provide $\sqrt{H}$ convergence of the time-varying parameter. Hence, assuming that the mapping in (4.7) is an ACM mapping and in addition

to some regularities conditions, the consistency and the asymptotic distribution of the NL-ILS estimator for the TVGARCH-in-mean model can be established by using the theory developed by Dominitz and Sherman (2005) and Giraitis, Kapetanios, and Yates (2010). The rates associated with the asymptotic results would, differently from the standard NL-ILS estimator adopted in Chapter 3, depend on the bandwidth parameter, H, and T. In this chapter, we do not show the asymptotic distribution of the kernel based NL-ILS estimator for the TVGARCH-in-mean specification, but we rely on the bootstrap framework (discussed in Section 4.3.1) to obtain the empirical distribution of the parameters governing the TVGARCH-in-mean model.

Chapter 3 shows that the NL-ILS estimator presents the additional feature to remain consistent even when the disturbances are no longer m.d.s processes, such as the cases of the weak-GARCH models in the spirit of Drost and Nijman (1993). This turns to be an important advantage of the NL-ILS estimator, since studies that estimate the risk premium function usually deal with daily, weekly or monthly data. These frequencies are obtained through the temporal aggregation of the observed intraday returns, which are a proxy for discretization of the continuous latent prices. Drost and Nijman (1993), Drost and Werker (1996) and Francq and Zakoian (2000) show that GARCH process are not closed under temporal aggregation, whereas weak-GARCH models are.

4.2.1 TVGARCH(1,1)-in-mean

We focus down our analyses on the TVGARCH(1,1)-in-mean specification, where we shall provide a step by step algorithm showing how to compute

the kernel based NL-ILS estimator. We also discuss the implementation of a bootstrap strategy to construct the confidence intervals for all the parameters θ.

We define the TVGARCH(1,1)-in-mean model as in (4.9) and (4.11).

$$y_t = \lambda_t \sigma_t + \epsilon_t \tag{4.9}$$

$$\epsilon_t = \sigma_t \eta_t \tag{4.10}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{4.11}$$

Similarly to the GARCH(1,1)-in-mean case, we assume that $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$ hold. Equation (4.11) allows an ARMA(1,1) representation as in (4.12), where $a = (\alpha + \beta)$ and $b = -\beta$.

$$\epsilon_t^2 = \omega + a\epsilon_{t-1}^2 + u_t + bu_{t-1} \tag{4.12}$$

Provided that $\alpha+\beta < 1$ holds, the AR polynomial in (4.12) can be inverted to generate an infinite MA process (MA(∞)) as:

$$\epsilon_t^2 = \psi_0 + \sum_{i=1}^{\infty} \psi_i u_{t-1} + u_t \tag{4.13}$$

where $\psi_0 = \frac{\omega}{1-a}$, $\psi_i = a^i(a+b)$. Denote $\phi = (\omega, a, b)'$, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_T)'$ and $\theta = (\lambda, \phi)'$. Identification of $\sum_{t=1}^{T} u_t^2$ with respect to the vector of parameters ϕ follows from Lemma 8 in Chapter 3. Following result (i) in Proposition 2, we use three kernel functions to retrieve estimates of λ_t. All kernels functions possess bounded first derivatives, however only the

Gaussian kernel has an infinite support.

$$K\left(\frac{t-\kappa}{H}\right) = \frac{1}{2} I\left(\left|\frac{t-\kappa}{H}\right| \leq 1\right), \quad \text{flat kernel} \tag{4.14}$$

$$K\left(\frac{t-\kappa}{H}\right) = \frac{3}{4}\left(1 - \left(\frac{t-\kappa}{H}\right)^2\right) I\left(\left|\frac{t-\kappa}{H}\right| \leq 1\right), \tag{4.15}$$

Epanechnikov kernel

$$K\left(\frac{t-\kappa}{H}\right) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{\frac{\left(\frac{t-\kappa}{H}\right)^2}{2}}, \quad \text{Gaussian kernel} \tag{4.16}$$

We are now in the position to discuss the implementation of the kernel based NL-ILS estimator, as well as feasible inference procedure. To this purpose, Subsection 4.2.1 displays the step by step procedure to compute the NL-ILS estimator, whereas Subsection 4.2.1 covers two different bootstrap strategies adopted to compute the confidence intervals associated with estimates of θ.

Kernel based NL-ILS algorithm

We compute the NL-ILS algorithm through the following steps:

Step 1: Choose an initial estimate of θ, such that $\widehat{\theta}_0 \in \mathbb{B}$, where $\mathbb{B}$ is the set of parameters satisfying the second-order stationarity conditions of (4.11)[3]. Applying $\widehat{\theta}_0$ to (4.9), (4.11) and (4.12), compute recursively estimates of the conditional variance, denoted as $\widehat{\sigma}^2_{0,t}$, and estimates of u_t, denoted by $\widehat{u}_{0,t}$.

Step 2: Using result in Proposition 2, compute $\widehat{\lambda}_{1,t}$ using any of the three

[3]In practise, define σ_y as the unconditional variance of y_t. Then, fix $\widehat{\lambda}_{0,t} = \left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\sigma_y y_\kappa\right][\sigma_y]^{-1}$ and obtain $\widehat{\epsilon}_{0,t}$. As a second step, estimate an AR(p) model having $\widehat{\epsilon}^2_{0,t}$ as dependent variable to obtain initial estimates of u_t. Finally, compute $\widehat{\phi}_0$.

kernels defined in (4.14), ,(4.15) and (4.16).

$$\widehat{\lambda}_{1,t} = \left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}^2_{0,\kappa}\right]^{-1} \sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}_{0,\kappa}y_\kappa \tag{4.17}$$

Step 3: Compute $\widehat{\epsilon}_{1,t}$ as $\widehat{\epsilon}_{1,t} = y_t - \widehat{\lambda}_{1,t}\widehat{\sigma}_{0,t}$. Using result (ii) in Proposition 2, obtain $\widehat{\phi}_1$ by minimizing the MA(∞) representation of the conditional variance using:

$$\widehat{\phi}_1 = \min_{\widehat{\phi}_1} \sum_{t=1}^{T} \left[\left[y_t - \widehat{\lambda}_{1,t}\widehat{\sigma}_{0,t}\right]^2 - \widehat{\psi}_{1,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{1,i}\widehat{u}_{0,t-1-i}\right]^2 \tag{4.18}$$

Step 4: Using $\widehat{\theta}_1$, compute recursively $\widehat{\sigma}^2_{1,t}$, $\widehat{\epsilon}_{1,t}$ and $\widehat{u}_{1,t}$ through (4.11), (4.9) and (4.12).

Repeat Steps 2, 3 and 4 j times until $\widehat{\theta}_j$ converges. Convergence occurs when the following both criteria are satisfied: $\left\|\widehat{\lambda}_{j,t} - \widehat{\lambda}_{j-1,t}\right\| \leq 10^{-5}$ and $\left\|\widehat{\phi}_j - \widehat{\phi}_{j-1}\right\| \leq 10^{-5}$. Note that the convergence bound is exogenously defined, making it possible to be as narrow as desired. Parameters on the j^{th} iteration are therefore given by:

$$\widehat{\lambda}_{j,t} = \left[\sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}^2_{j-1,\kappa}\right]^{-1} \sum_{\kappa=1}^{T} K\left(\frac{t-\kappa}{H}\right)\widehat{\sigma}_{j-1,\kappa}y_\kappa \tag{4.19}$$

$$\widehat{\phi}_j = \min_{\widehat{\phi}_j} \sum_{t=1}^{T} \left[\left[y_t - \widehat{\lambda}_{j,t}\widehat{\sigma}_{j-1,t}\right]^2 - \widehat{\psi}_{j,0} - \sum_{i=0}^{\bar{q}} \widehat{\psi}_{j,i}\widehat{u}_{j-1,t-1-i}\right]^2 \tag{4.20}$$

Bootstrap algorithm

We perform inference on the NL-ILS estimator by using the bootstrap framework. We adopt two distinct strategies: the first one is the full parametric bootstrap, whereas the second one follows the wild bootstrap pro-

posed in Linton and Perron (2003). As discussed in Section 4.2, provided that the maximum eigenvalue of $V(\theta)$ is strictly smaller than one in absolute value, $\sqrt{T}\left(\widehat{N}_T(\theta) - \theta\right) \xrightarrow{d} N(0, \Sigma)$, and some additional regularities conditions, the theory developed by Dominitz and Sherman (2005) guarantees that the NL-ILS estimator is asymptotically normally distributed. This strengthens the bootstrap validation we adopt in this chapter, since under high-level assumptions the kernel based NL-ILS estimator is asymptotically well behaved. The parametric and the wild bootstrap differ only in the first step. To all the remaining steps, we do not differentiate from the two algorithms.

1. Given the NL-ILS estimates $\widehat{\lambda}_t$ for $t = 1, 2, ..., T$, $\widehat{\omega}$, $\widehat{\alpha}$ and $\widehat{\beta}$, compute the recentered residuals $\widehat{\epsilon}_t^c$, such that $\widehat{\epsilon}_t^c = \widehat{\epsilon}_t - \bar{\widehat{\epsilon}}_t$ and $\bar{\widehat{\epsilon}}_t = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_t$.

   i. Parametric: Bootstrap $\widehat{\epsilon}_t^c$ to generate a $(T \times 1)$ vector of residuals denoted by ϵ_t^b.

   ii. Wild: As in Linton and Perron (2003), define z_t as a variable with $\mathbb{E}\left(z_t^j\right) = 0$ for $j = 1, 3, ...$ and $\mathbb{E}\left(z_t^j\right) = 1$ for $j = 2, 4, ...$ Similarly to them, we set $z_t = 1$ or $z_t = -1$ with probability equal to 0.5. Generate $\epsilon_t^b = \epsilon_t^c z_t$.

2. Set σ_1^2 and ϵ_1^c as starting values. Using ϵ_t^b, $\widehat{\lambda} = \left(\widehat{\lambda}_1, ..., \widehat{\lambda}_T\right)'$ and $\widehat{\phi} = \left(\widehat{\omega}, \widehat{\alpha}, \widehat{\beta}\right)'$, compute bootstrapped values of y_t, denoted by y_t^b.

3. Using $\left\{y_t^b\right\}_{t=1}^{T}$, estimate $\widehat{\lambda}^b = \left(\widehat{\lambda}_1^b, ..., \widehat{\lambda}_T^b\right)'$ and $\widehat{\phi}^b = \left(\widehat{\omega}^b, \widehat{\alpha}^b, \widehat{\beta}^b\right)'$ by adopting the kernel based NL-ILS estimator as discussed in 4.2.1.

4. Repeat steps 1, 2 and 3 B times[4].

[4]We set B = 1000 in both empirical and Monte Carlo studies.

5. Compute the percentiles and standard deviation from the empirical distribution of $\widehat{\lambda}^b = \left(\widehat{\lambda}_1^b, ..., \widehat{\lambda}_T^b\right)'$ and $\widehat{\phi}^b = \left(\widehat{\omega}^b, \widehat{\alpha}^b, \widehat{\beta}^b\right)'$ where $b = 1, 2, ..., B$..

Both bootstrap procedures described above are highly time-demanding[5], which makes a proper coverage probability study based on Monte Carlo validation very difficult to be undertaken. We discuss the coverage probability associated with the two methodologies in Subsection 4.3.1.

4.3 Numerical Illustrations

4.3.1 Monte Carlo

This section has manly two objectives. Firstly, we assess the performance of the kernel based NL-ILS estimator on estimating the parameters governing the TVGARCH(1,1)-in-mean model. Secondly, we discuss how the two bootstrap methodologies discussed in Section 4.2.1 perform on retrieving the confidence intervals associated with the NL-ILS estimates of λ_t.

Regarding the first point, we focus on understanding how the NL-ILS estimator tracks the time-varying risk premium parameters in terms of the root mean squared error (RMSE) and point-wise correlation with the latent time-varying coefficients. To this purpose, we implement a variety of bandwidth choices (different degrees of smoothing) that will be very useful to guide our choice of H when estimating the time-varying risk premium parameters in Subsection 4.3.2. From the nonparametric literature, there is a variance-bias tradeoff involving the choice of the bandwidth parameter

[5]Computing the NL-ILS estimates confidence intervals for a TVGARCH(1,1)-in-mean model (using only one kernel function) with $T = 2000$ takes one day in a dedicated server (one core).

H. In one hand, if H is too small, bias associated with the kernel based NL-ILS estimates tend to decrease, whereas their variance increases. On the other hand, if H is too large, bias increases and variance decreases. In fact, the choice of H turns to be more important than choosing the kernel function. From Giraitis, Kapetanios, and Yates (2010), $H = T^{0.5}$ is the closest value to the optimal H that minimizes the mean squared error (MSE) in their time-varying AR(1) model. As discussed in Section 4.2, we require the bandwidth parameter to satisfy the following: $H \longrightarrow \infty$ and $H(T) = o(T)$. We evaluate the performance of the kernel based NL-ILS estimator under the following bandwidth parameters: $H = T^{0.2}, H = T^{0.3}, H = T^{0.4}, H = T^{0.5}, H = T^{0.6}, H = T^{0.7}$ and $H = T^{0.8}$. Finally, we also evaluate the finite sample performance of the NL-ILS estimator when estimating parameters in the conditional variance equation.

We specify two different data generation processes in this subsection. Both models are TVGARCH(1,1)-in-mean models as depicted in (4.9) and (4.11). We set η_t to be normally distributed with zero mean and variance equals to one. The difference between the two models consists on the specification of the time-varying risk premium parameter λ_t. In the first case, we define λ_t as a bounded random walk process as in (4.5), with $a_T = a_{t-1} + \xi_t$ and $c = 0.9$. We add some dependence on ξ_t, by modelling it as an AR(1) process with $\xi_t = \rho\xi_{t1} + \varsigma_t$, where $\varsigma_t \sim N(0, 0.02)$. The second specification sets λ_t as a covariance stationary process. We specify λ_t as an AR(1) process with autoregressive parameter equal to 0.9. This second specification is not supported by the theory developed in Giraitis, Kapetanios, and Yates (2010) (see Remark 2.4), however it sheds light about the performance of the kernel based NL-ILS estimator when λ_t is

not persistent enough[6]. In all exercises, we fix the number of replications to 1000 unless otherwise stated. We also discard the initial 500 observations to reduce dependence on initial conditions. All models are estimated using the CML[7] optimization library in GAUSS.

Table 4.2 displays the results associated with the first specification, where λ_t is modeled as a bounded random walk process. Considering the fit of the kernel based NL-ILS estimator, we conclude that the best choices for bandwidth parameters, the ones that minimize the RMSE, are either $T^{0.5}$ or $T^{0.6}$. Furthermore, these are also the bandwidths which deliver the highest point-wise correlation (around 0.85) between the NL-ILS estimates and the true latent time-varying risk premium parameter. Figures 4.1, 4.2,4.3, 4.4, 4.5 and 4.6 display the evolution of λ_t, $\widehat{\lambda}_t$ and its correspondent confidence intervals. From these figures, we assert that the kernel based NL-ILS estimator provides estimates (considering all the three alternative kernel functions) that track λ_t very accurately, corroborating the point-wise correlation result. Considering the performance of the kernel based NL-ILS estimator on recovering the parameters in the conditional variance equation, we conclude that apart from the scenario where $H = T^{0.2}$, all different combinations of kernel methods and bandwidth parameters deliver unbiased estimates of $\phi = (\omega, \alpha, \beta)'$. It also relevant to point out that the RMSE of ϕ is reasonably small and constant through all the different kernel

[6]Robinson (1989) and Orbe, Ferreira, and Rodriguez-Poo (2005) establish the consistency of kernel based OLS estimators when dealing with regressions that present deterministic time-varying coefficients. They impose smoothness assumptions on λ_t to obtain consistency.

[7]CML (Constrained Maximum Likelihood Estimation) is library in GAUSS designed to solve maximum likelihood functions subject to linear and nonlinear constraints. In all Monte Carlo simulations, we set global variables in CML to their default values, because this specification is flexible enough to accommodate endogenous changes in both algorithms and grid search procedures.

functions and bandwidth choices, indicating that NL-ILS estimates of ϕ are robust to different bandwidth and kernel choices.

Table 4.3 reports results considering the case where λ_t is an AR(1) process, such that $\lambda_t = \rho\lambda_{t-1} + \varsigma_t$, with $\rho = 0.9$ and $\varsigma_t \sim N(0, 0.2)$. We find that the kernel based NL-ILS estimator looses performance in terms of RMSE and point-wise correlation in all different scenarios. This turns not to be a surprising result, because λ_t is a now covariance stationary process and kernel methods cannot handle such feature (see discussion in Giraitis, Kapetanios, and Yates (2010)). To capture the less persistent feature of λ_t, $H = T^{0.3}$ turns to be the best bandwidth choice considering the RMSE and point-wise correlation tradeoff. We stress, however, that even when λ_t is a covariance stationary process, the kernel based NL-ILS estimator delivers unbiased estimates of the parameters in the conditional variance equation.

We now turn our attention to the performance of the two bootstrap methodologies discussed in Section 4.2.1. To give a flavour about the coverage probability of these two strategies, we compute the coverage probability associated with different confidence intervals (CI) for one realization of the TVGARCH(1,1)-in-mean model. To assess the magnitude of the point-wise confidence bands, we compute the root mean squared distance (RMSD) between the point-wise upper and the lower bound associated with different confidence intervals adopted (90%, 95% and 99%), such that $RMSD = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\lambda}_t^u - \widehat{\lambda}_t^l\right)^2\right]^{0.5}$, where $\widehat{\lambda}_t^u$ and $\widehat{\lambda}_t^l$ account for the point-wise upper and lower bound associated with a specific confidence interval. Table 4.1 shows that both the parametric and the wild bootstrap perform reasonable well, delivering coverage probabilities very close to the theoretical values implied by the confidence interval. This strengthens our claim

that these two bootstrap methodologies qualify as an inference tool for constructing the confidence bands associated with the kernel based NL-ILS estimator. Regarding the magnitude of the confidence intervals, we report an average distance of 0.38 when $CI = 90\%$, which may be too high when dealing with empirical applications. In fact, the high values of RMSD reported in both bootstrap methodologies reenforce the difficulties associated with estimating time-varying parameters in the presence of latent regressors, as it is the case of the TVGARCH-in-mean models. Hence, as discussed previously, although the kernel based NL-ILS estimator has wide confidence bands, it tracks the dynamics of the time-varying parameter λ_t very well, providing an important insight on the behaviour of the time-varying risk premium parameter.

4.3.2 Empirical results

We examine the time-varying risk premium parameter using the TVGARCH(1,1)-in-mean framework. We estimate λ_t using the NL-ILS estimator as discussed in Subsection 4.2.1. We adopt the excess returns computed using the CRSP value-weighted index aggregated on weekly and monthly basis[8]. We choose the CRSP index because it is considered the financial index that best mimics the entire market, including large and small capitalized firms. Moreover, Chapter 3 documents that risk premium parameter λ obtained through a GARCH(1,1)-in-mean model estimated with the robust NL-ILS estimator is statistically significant only when the CRSP index is adopted. We show that for less complete indices,

[8]We obtain the market excess returns through Wharton Research Data Services (wrds), Fama French & Liquidity Factors library. This variable is denoted as MKTRF on wrds database and it is available on daily and monthly basis.

such as the S&P100 and S&P500 indices, estimates of λ, are not statistically significant, indicating that market coverage plays an important role on identifying the risk premium parameter. We work with two different sampling frequencies: weekly and monthly, yielding 2,426 and 1,023 observations, respectively. Monthly data is available since 1926, which give us the opportunity to cover the Great Depression and the financial crisis of 2007/08. Table 4.4 displays the descriptive statistics.

Figures 4.7 and 4.8 plot weekly estimates of λ_t considering $H = T^{0.5}$ and $H = T^{0.6}$, respectively. We choose these two bandwidth values, because they present the best performance in terms of RMSE in the Monte Carlo study discussed in Section 4.3.1. We also plot the 90% upper and lower confidence intervals computed using the empirical percentiles obtained through the parametric bootstrap[9]. With respect to the point-wise analyses, we find that there is strong evidence that the risk premium parameter is indeed time-varying, with $\widehat{\lambda}_t$ assuming both positive and negative values, within ranges of $(0.25, -0.25)$ and $(0.2, -0.1)$ for the $H = T^{0.5}$ and $H = T^{0.6}$, respectively. This result reinforces the claim that specifying the risk premium parameter as time-invariant or as a deterministic function of the conditional standard deviation can cause severe bias on the estimates of the risk premium function. Furthermore, periods of negative risk-return tradeoff can arise as part of the volatility feedback mechanism, as pointed out by Campbell and Hentschel (1992) and Dahl and Iglesias (2009). In fact, periods of financial distress usually present high volatility, which leads to an increase in the risk premium and the discount rate. These cause a drop in prices, yielding to a momentaneous negative relationship between

[9]Estimates of λ_t computed with $H = T^{0.2}$, $H = T^{0.3}$ and $H = T^{0.4}$ are available upon request.

volatility and returns. This is indeed the picture we find on Figures 4.7 and 4.8, where $\widehat{\lambda}_t$ turns negative for short periods of time.

Analysing 4.7 more in depth, we find that $\widehat{\lambda}_t$ is very volatile when the Epanechnikov and flat kernels are adopted. This contributes for the bootstrap confidence bands to be very wide, which tends to jeopardize the significance analyses. We find that $\widehat{\lambda}_t$ estimated with these two kernel methods are statistically different from zero for the period prior to the year of 2000. This is exactly the period that precedes the Dot-com bubble, which lead to an eight-month recession starting in March 2001 and lasting until November 2001.

The picture described above is even clearer when considering Figure 4.8, where the bandwidth is set equal to $T^{0.6}$. This leads to much smoother estimates of λ_t, making both point-wise and significance analyses more relevant. We find that $\widehat{\lambda}_t$ does present a strong variation over time, picking in periods prior to financial distress. In fact, considering the period covering the last twenty years (1991 - 2011), we find that estimates of the time-varying risk premium during this time frame present a cyclical pattern, picking in periods that precede the financial crises. Also, the periods where $\widehat{\lambda}_t$ is negative or approaches to zero coincide with the intervals of time which the economy is going through a recession. This indicates that the volatility feedback mechanism takes action, leading to a drop in stock prices, which usually anticipates business cycles fluctuations. Therefore, we assert that $\widehat{\lambda}_t$ estimated using the TVGARCH(1,1)-in-mean framework is able to track both *bear market* and business cycle expansions and contractions. To be more precise in our analysis, we focus on the first plot of Figure 4.8. Considering the NL-ILS estimator computed with the Epanechnikov

kernel, we find that $\widehat{\lambda}_t$ picks on the week of 14/Jun/1996, which precedes the Russian crisis. From this date onwards, $\widehat{\lambda}_t$ declines becoming negative on the week of 02/Feb/2001. This time range coincides with the recession period reported by National Bureau of Economic Research (NBER), that states that the United States (US) economy was in recession during the period starting on the March 2001 until November 2001 (see Table 4.5 for the specific dates). The time-varying risk premium parameter presents a similar pattern in the last half of the first decade of the twenty first century. We find that $\widehat{\lambda}_t$ picks on the week of 18/Feb/2005, starting a downturn that results in negative values associated with $\widehat{\lambda}_t$ on the week of 17/Nov/2006. The time-varying risk premium only turns positive on the week of 26/Sep/2008. This pattern again tracks and anticipates both the *bear market* and the US recession dates. Regarding the latter, the NBER reports that the US economy faced recession from Dec/2007 until Jun/2009. This again provides us with a date intersection between the behaviour of the real economy and the time-varying risk premium parameter.

With regard to tracking the *bear market* period, we find that the downturn of $\widehat{\lambda}_t$ coincides with the period prior to the failure of the Lehman Brothers (13/Sep/2008), including the burst of the housing bubble and the bailout of a series of financial institutions including the Northern Rock, Fannie Mae, Freddie Mac, American International Group (AIG) among others. Figure 4.9 plots $\widehat{\lambda}_t$ and its confidence bands together with the conditional standard deviation computed using the TVGARCH(1,1)-in-mean specification. We find that in periods where $\widehat{\lambda}_t$ is high, market volatility is low. When $\widehat{\lambda}_t$ is either negative or presents a declining path, we observe the volatility associated with the excess returns is very high. These corrob-

orate our claim that $\widehat{\lambda}_t$ is able to track both financial market performance and business cycles fluctuations.

We now turn our attention to the monthly estimates of $\widehat{\lambda}_t$. Figures 4.10 and 4.11 display plots considering the two alternative bandwidth choices adopted in this section: $H = T^{0.5}$ and $H = T^{0.6}$, respectively. The monthly sample carries an important difference from the one at weekly bases we discussed previously: it spans from a longer period (1926-2011), comprehending events such as the Great depression, the World War II and the post-war period. Moreover, this much longer sample also allows us to investigate potential changes on the time-varying risk premium parameter behaviour, following potential structural changes in the economy. In fact, these structural changes may arise from a wide variety of factors, including changes in the investors preferences, financial markets organization, market regulation, portfolio composition and availability of assets. Comparing the results obtained with weekly and monthly frequencies, we expect a trade-off between smoothness of $\widehat{\lambda}_t$ and the responsiveness of the time-varying risk premium parameter to shocks on the CRSP index. As a consequence of that, estimates of λ_t computed using monthly data are smoother and more persistent that the estimates we report in Figures 4.7 and 4.8. As a drawback, we have that monthly estimates of λ_t tend to lose power on predicting financial crises and economic recessions when compared to their weekly counterparts.

Considering the set of graphs where $H = T^{0.5}$ and focusing on the estimates computed using the Epanechnikov and flat kernels, we find that $\widehat{\lambda}_t$ is statistically different from zero in the period that precedes the year of 2000 (from Apr/1994 to Apr/1998 for the Epanechnikov kernel, and from

Feb/1994 to Dec/1998 for the flat kernel). This finding corroborates our previous results considering weekly estimates of λ_t computed using $H = T^{0.6}$, where $\widehat{\lambda}_t$ turns to be significant from the week of 22/Jul/1998 until the week of 17/Jul/1998. Furthermore, we find that $\widehat{\lambda}_t$ is significantly different from zero within the Nov/1940 - Mar/1964 and Jun/1939 and Dec/1964 for the Epanechnikov and flat kernel functions, respectively. These periods of significant parameters are by far greater than the ones observed during the nineties, suggesting a structural change on the pattern of the time-varying risk premium parameter.

As in the weekly frequency analyses, we focus our analyses on the estimates of λ_t computed using $H = T^{0.6}$ and two different kernel specifications: the Epanechnikov and flat kernels. These choices are supported by our Monte Carlo results, that indicate that the $H = T^{0.6}$ is the bandwidth choice that minimizes the RMSE for the Epanechnikov and flat kernel functions. The first and third plot of Figure 4.11 display $\widehat{\lambda}_t$ computed with the Epanechnikov and flat kernel functions, respectively. We find that under these specifications the confidence bands are narrower than the ones computed with $H = T^{0.5}$. We find $\widehat{\lambda}_t$ is statistically significant in 46.5% and 43.0% of the total observations. These are extremely interesting results, because they shed light on the mixed evidence reported in the literature regarding sign and significance of the risk premium parameter. From the results in Figure 4.11, we find that following the persistent time-varying nature of λ_t, it is misleading to model the risk premium parameter as a time-invariant parameter. In other words, if we model λ_t as a time-invariant parameter, we are likely to obtain results that falsely return insignificant or barely significant estimates of the risk premium parameter.

We now focus on the relation between the time-varying risk premium parameter and periods of financial distress and business cycle fluctuations. Figure 4.12 displays plots of $\widehat{\lambda}_t$ with its respective upper and lower 90% confidence bands and the conditional standard deviation computed using the TVGARCH(1,1)-in-mean specification. As in the previous analyses, we focus on the estimates obtained with the Epanechnikov and flat kernel functions. We find that during the first period where λ_t is statistically different from zero (1939/40 - 1964), the US economy faces five periods of recessions, however the time-varying risk premium parameter is very high. When analysing the second period where $\widehat{\lambda}_t$ is significantly different from zero (May/1982-Jul/1998 and Mar/1988 - Apr/1998, Epanechnikov and flat kernel functions, respectively) we find that there is only one period where the US economy faces recession (Jul/1990-Mar/1991). Considering the next two recessions, (Mar/2001-Nov/2001 and Dec/2007-Jun/209), we find that $\widehat{\lambda}_t$ is not statistically different from zero, being in fact lower than 0.1. Differently from the results obtained using weekly data, we do not observe a spike on $\widehat{\lambda}_t$ in the period that precedes the 2007-09 financial crisis, indicating that the boon on equity prices observed from 2001 to 2007 was in fact associated with low values of $\widehat{\lambda}_t$. To conclude, we find strong evidences that the risk premium parameter is time-varying, and therefore needs to be modelled as so. Moreover, we find that the relation between the significance on the time-varying risk premium parameter and business cycle fluctuations change over time, suggesting that it has become weaker in the last twenty years.

Considering the relationship between λ_t and σ_t, Linton and Perron (2003), Christensen, Dahl, and Iglesias (2012) and Conrad and Mammen

(2008) relax the linearity assumption governing the risk premium function, finding that the risk premium function exhibits a hump shape. In this chapter, however, we force the relation between λ_t and σ_t to be linear, but we allow λ_t to evolve stochastically. This implies that the relation between λ_t and σ_t in no longer deterministic in our approach. Figures 4.13 and 4.14 depict scatter plots of $\widehat{\lambda}_t$ versus $\log(\widehat{\sigma}_t^2)$[10] and the risk premium function versus $\log(\widehat{\sigma}_t^2)$. Our aim is to investigate whether there is a clear relation between these variables. We find that such relations do not hold under the TVGARCH(1,1)-in-mean framework. Regarding the first series of graphs ($\widehat{\lambda}_t$ versus $\log(\widehat{\sigma}_t^2)$), we cannot identify any pattern. We can only say that when the volatility increases to values above 0.04, $\widehat{\lambda}_t$ is either negative or below 0.05, indicating that volatility feedback mechanism is the key force driving the λ_t towards negative or zero values. Regarding the series of graphs displaying the relationship between the risk premium function and $\widehat{\lambda}_t$ and $\log(\widehat{\sigma}_t^2)$, we find that such relation is highly nonlinear, confirming that the TVGARCH(1,1)-in-mean specification is flexible enough to accommodate different shapes of the risk premium function.

4.4 Conclusion

In this chapter we model the risk-return tradeoff allowing for the risk premium parameter to be time-varying and evolve stochastically over time as a random walk process. To this purpose, we introduce the time-varying GARCH-in-mean (TVGARCH-in-mean) model. We introduce the kernel based NL-ILS estimator and show that it successfully estimates the time-

[10]We choose to construct the graphs using $\log(\widehat{\sigma}_t^2)$ in order to be in accordance with the notation used in Linton and Perron (2003).

varying risk premium parameter, λ_t. The kernel based NL-ILS estimator generalizes the kernel based OLS estimator implemented in Giraitis, Kapetanios, and Yates (2010), making it possible to estimate λ_t in the presence of a latent regressor (σ_t) under the TVGARCH-in-mean specification. The Monte Carlo study shows that the kernel based NL-ILS estimator presents a good finite sample performance on estimating both λ_t and the parameters in the conditional variance equation. Furthermore, we show that the parametric and wild bootstrap methodologies can be implemented to compute the confidence intervals associated with all parameters governing the TVGARCH-in-mean model.

We investigate the time-varying risk premium parameter using the excess returns computed using the CRSP value-weighted index aggregated on weekly and monthly basis. By adopting the TVGARCH-in-mean specification, we address the issue of misspecification of the conditional mean, as it is regarded as one of the causes for mixed evidences regarding the significance and sign of the risk premium parameter. Also, by relying on the robust NL-ILS estimator, we address the issue of biased results following misspecification in the conditional variance equation. We find strong evidences, on both sample frequencies, that λ_t is indeed time variant. Considering the monthly frequency, we find that estimates of λ_t are statistically different from zero in up to 46.5% of the observations. This result sheds light on the mixed evidences regarding sign and significance of the risk premium parameter, because modelling the risk premium parameter as a time-invariant coefficient may lead to biased results. Furthermore, we find that estimates of λ_t computed using the weekly excess returns track and anticipate both *bear market* phases and business cycles fluctuations. In par-

ticular, we find that periods of financial distress and economic recessions are preceded by downturn on the time-varying risk premium parameter, whereas during the financial crisis, the time-varying risk premium parameter is close to zero or even negative. Finally, our results suggest that the relation between significance of the time-varying risk premium parameter and business cycle fluctuations has changed in the past twenty years.

4.5 Appendix

Figure 4.1: Parametric Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - Epanechnikov kernel

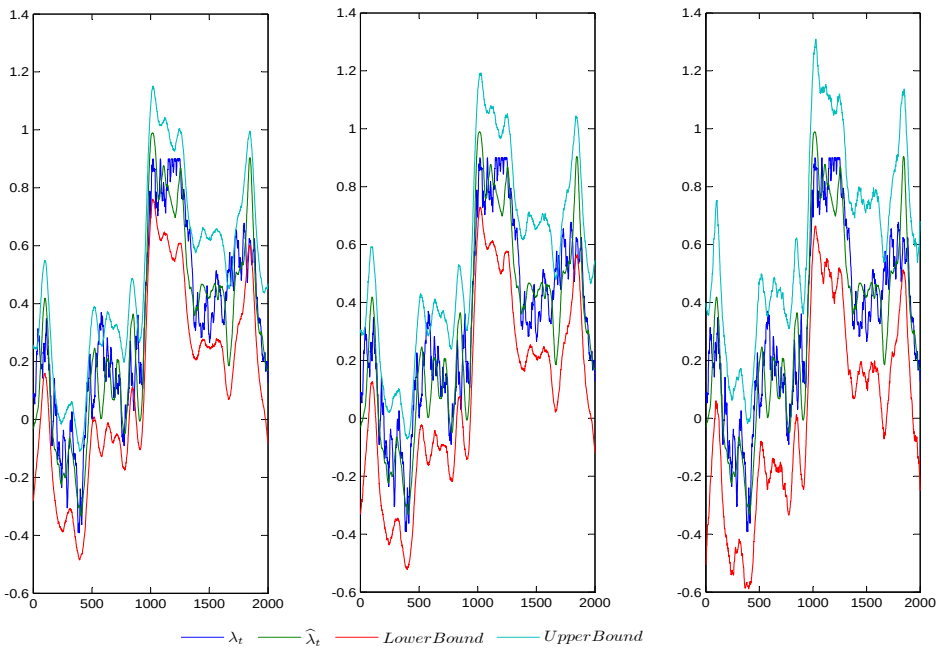


Figure 4.1 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the Epanechnikov kernel function as in (4.15).

Figure 4.2: Wild Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - Epanechnikov kernel

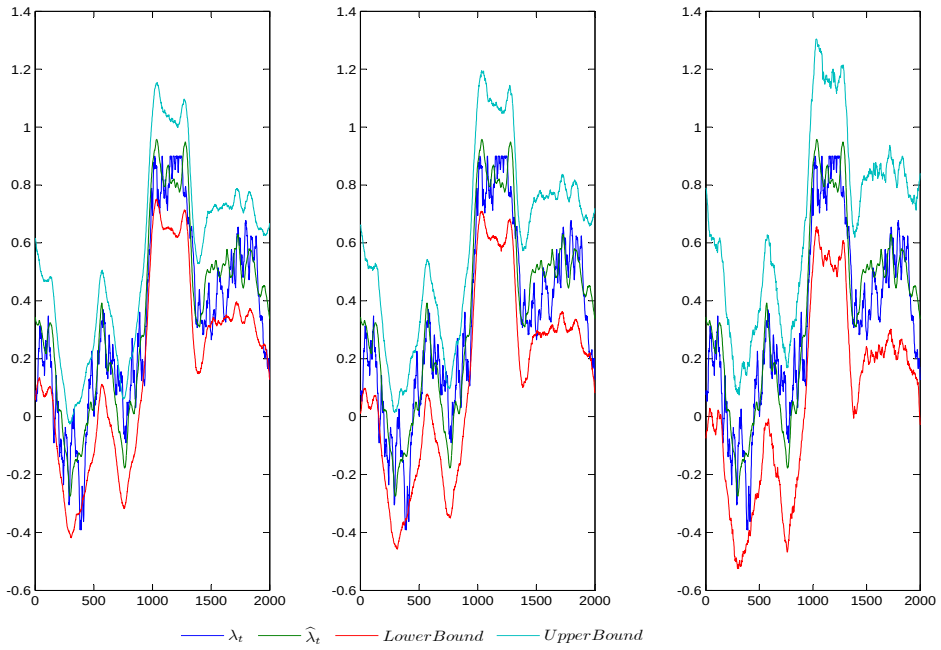


Figure 4.2 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the wild bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the Epanechnikov kernel function as in (4.15).

Figure 4.3: Parametric Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - Gaussian kernel

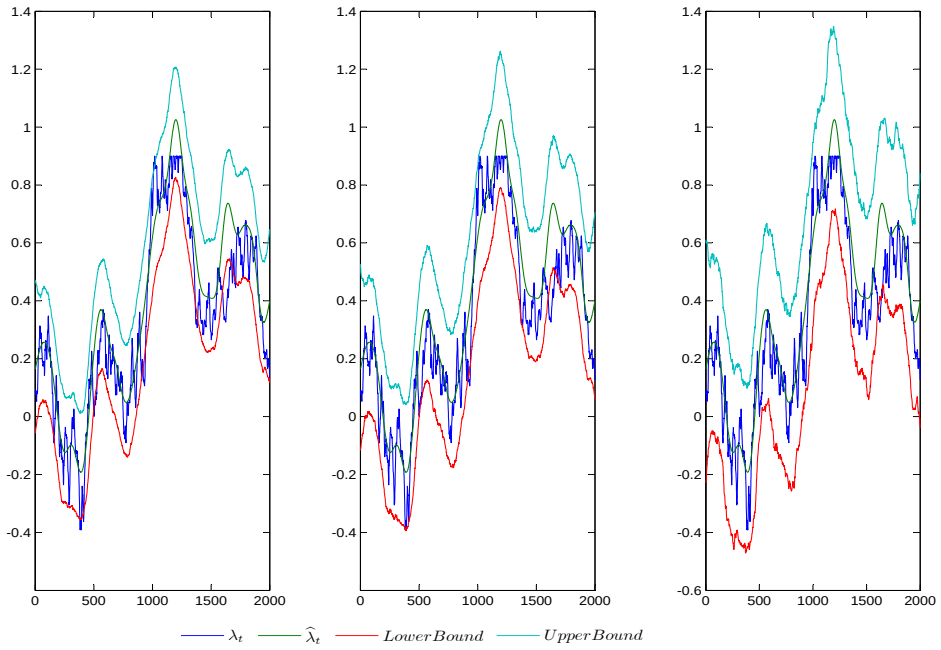


Figure 4.3 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the Gaussian kernel function as in (4.16).

Figure 4.4: Wild Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - Gaussian kernel

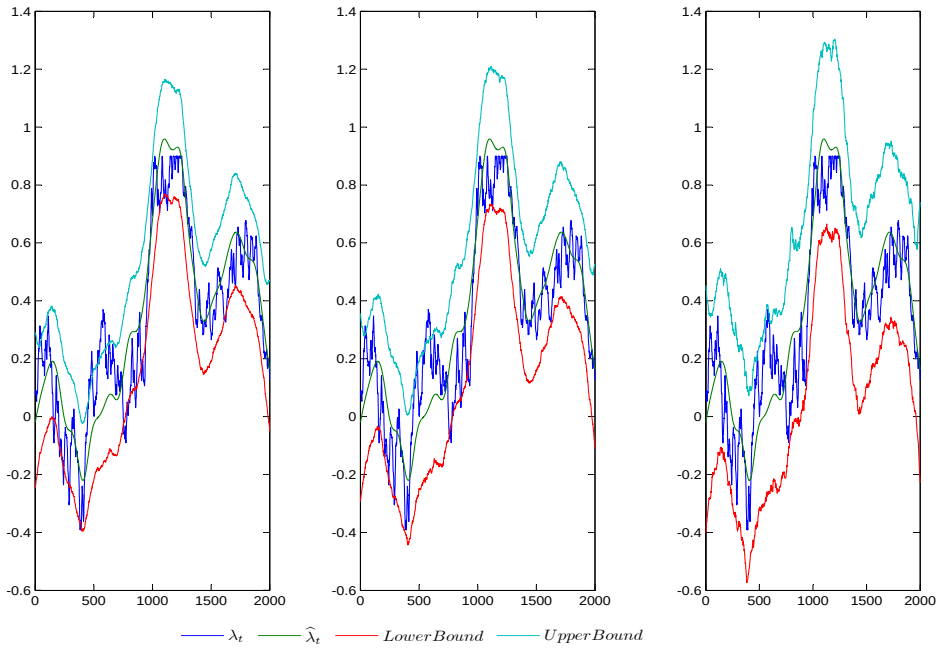


Figure 4.4 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the wild bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the Gaussian kernel function as in (4.16).

Figure 4.5: Parametric Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - flat kernel

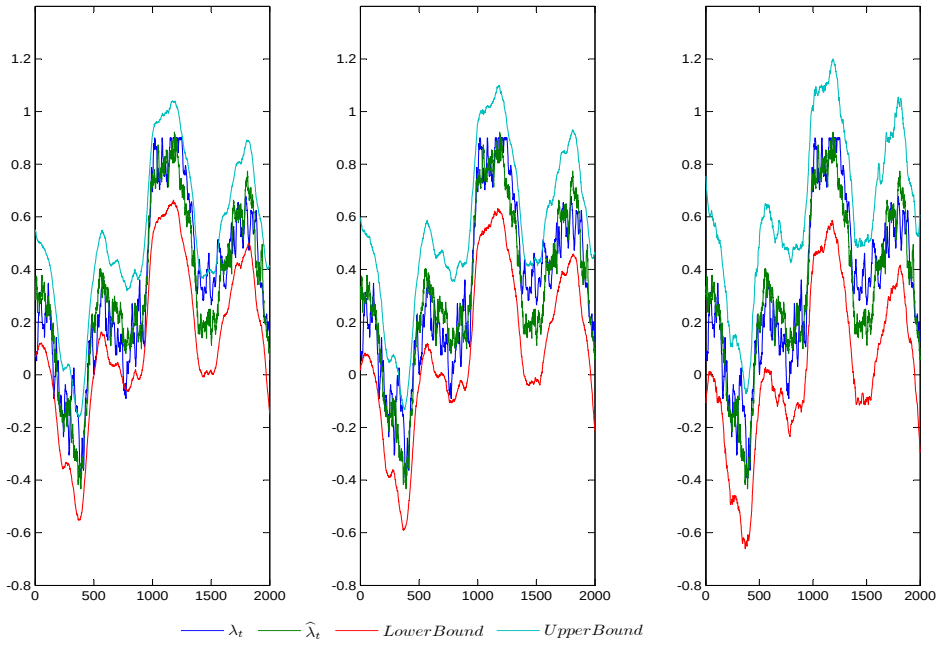


Figure 4.5 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the flat kernel function as in (4.14).

Figure 4.6: Wild Bootstrap Confidence Intervals - TVGARCH(1,1)-in-mean - flat kernel

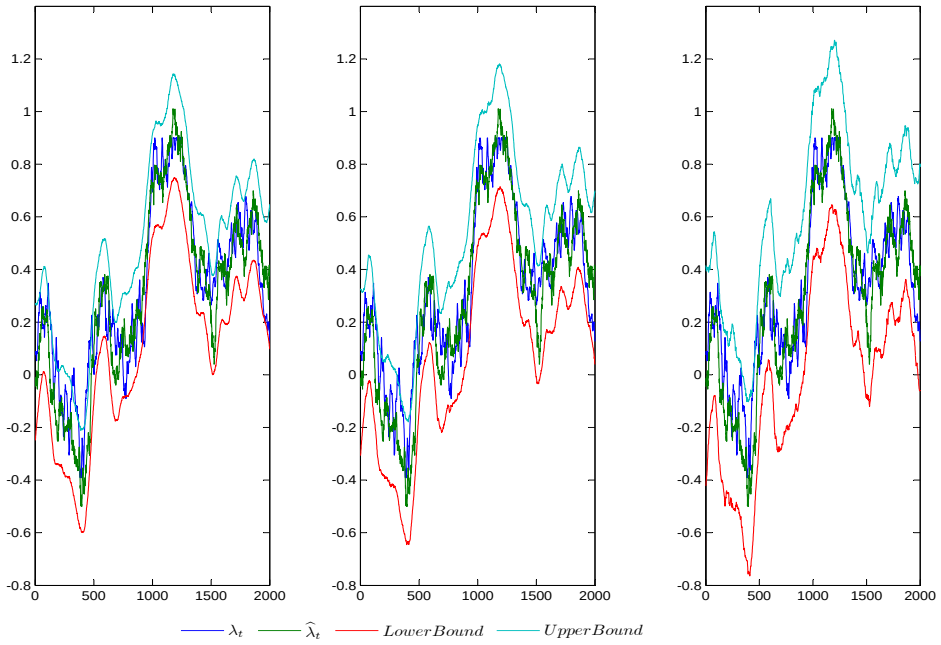


Figure 4.6 plots confidence intervals (90%, 95% and 99%) computed using the empirical percentiles obtained with the wild bootstrap discussed in Section 4.2. The plot on the left hand side depicts the 90% confidence interval, whereas the graphs on the center and on the right hand side display the 95% and 99% confidence intervals, respectively. We generate the TVGARCH(1,1)-in-mean model defining λ_t as in (4.5) and setting the parameters in the conditional variance equation as $\omega = 0.01$, $\alpha = 0.05$ and $\beta = 0.90$. We perform 1000 replications in the bootstrap algorithm. Estimates of the time-varying risk premium parameters, $\widehat{\lambda}_t$, are computed using the NL-ILS estimator computed with the flat kernel function as in (4.14).

Figure 4.7: Time-varying risk premium estimation - weekly data

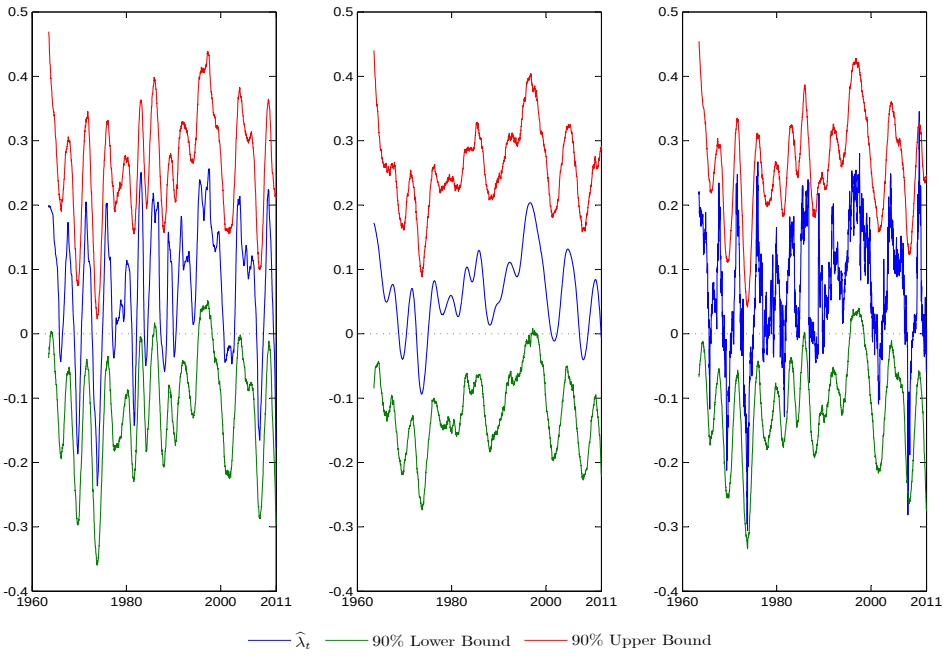


Figure 4.7 plots estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.5}$.

Figure 4.8: Time-varying risk premium estimation - weekly data

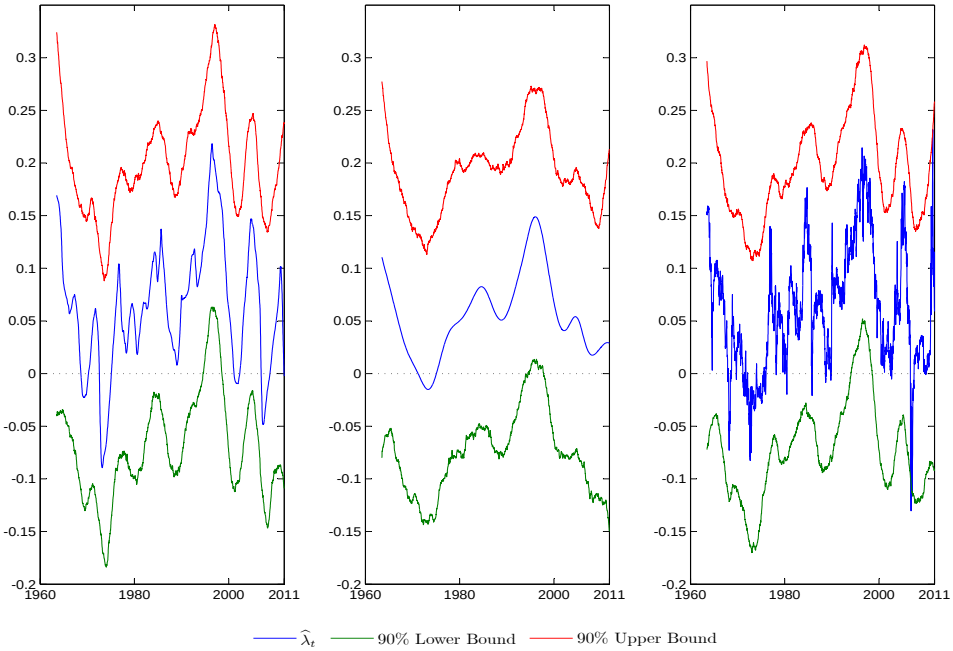


Figure 4.8 plots estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.6}$.

Figure 4.9: Time-varying risk premium estimation and conditional standard deviation - weekly data

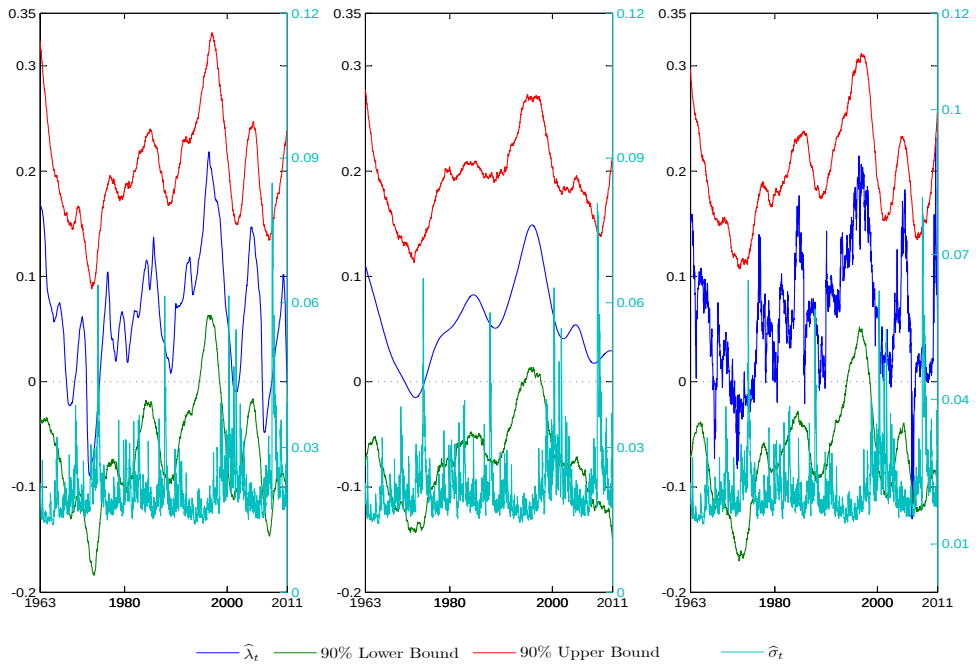


Figure 4.9 plots, on the left axis, estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. On the right axis, we plot, in light blue, the conditional standard deviation computed using the TVGARCH(1,1)-in-mean specification. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.6}$.

Figure 4.10: Time-varying risk premium estimation - monthly data

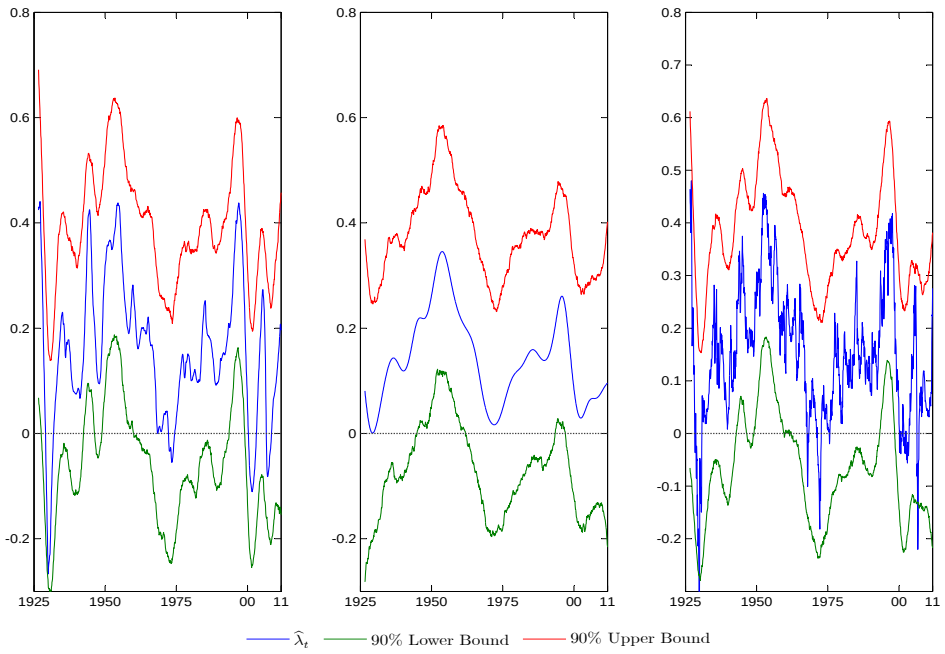


Figure 4.10 plots estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.5}$.

Figure 4.11: Time-varying risk premium estimation - monthly data

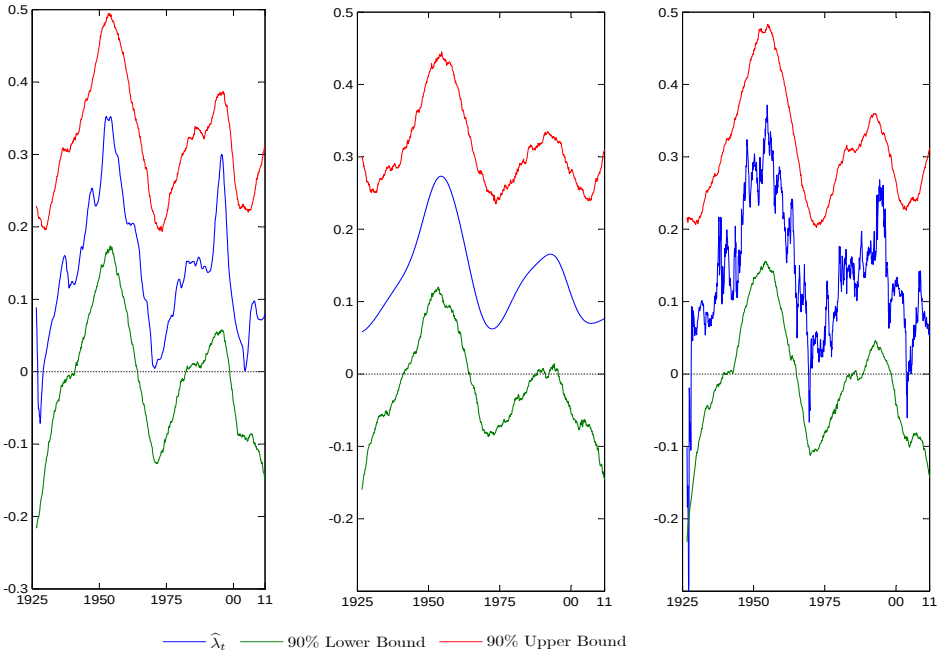


Figure 4.11 plots estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.6}$.

Figure 4.12: Time-varying risk premium estimation and conditional standard deviation - monthly data

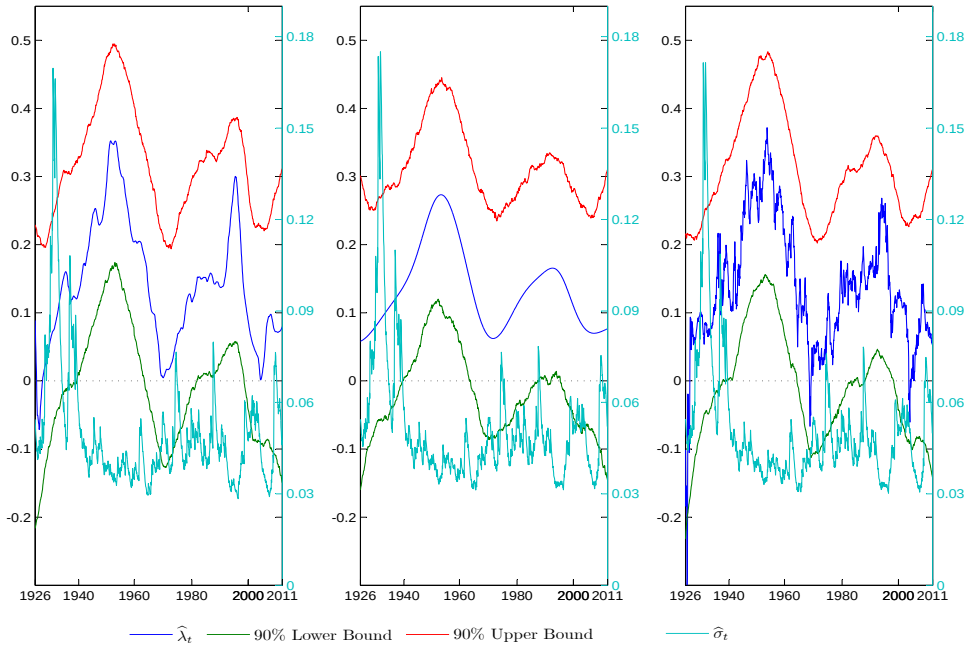
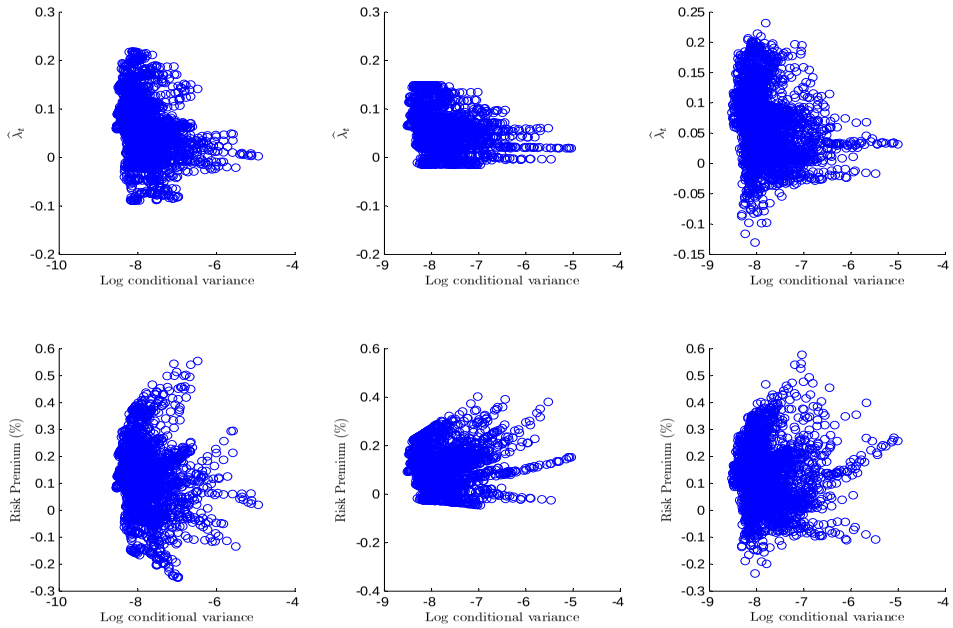


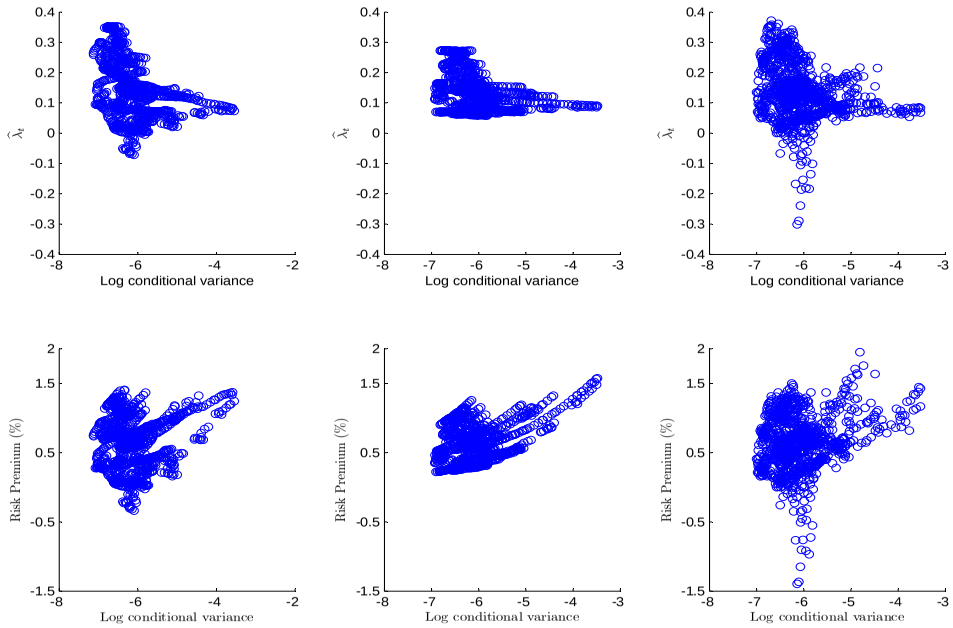
Figure 4.12 plots, on the left axis, estimates of λ_t and the 90% confidence intervals computed using the empirical percentiles obtained with the parametric bootstrap discussed in Section 4.2.1. On the right axis, we plot, in light blue, the conditional standard deviation computed using the TVGARCH(1,1)-in-mean specification. We perform 1000 replications in the bootstrap algorithm. The plot on the left hand side depicts estimates of the time-varying risk premium parameters computed using the NL-ILS estimator computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display estimates of λ_t computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively. We fix the bandwidth parameter equal to $T^{0.6}$.

Figure 4.13: $\widehat{\lambda}_t$ and risk premium (%) versus log conditional variance - weekly data



Plots on the first row of Figure 4.13 displays $\widehat{\lambda}_t$ against $ln\left(\widehat{\sigma}_t^2\right)$ computed using estimates of the TVGARCH(1,1)-in-mean model. Graphs on the second row of Figure 4.13 plots the risk premium, $\widehat{\lambda}_t\widehat{\sigma}_t$, versus $ln\left(\widehat{\sigma}_t^2\right)$. The plots on the left hand side are constructed using the NL-ILS estimator computed computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display scatter plots computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively.

Figure 4.14: $\widehat{\lambda}_t$ and risk premium (%) versus log conditional variance - monthly data



Plots on the first row of Figure 4.14 displays $\widehat{\lambda}_t$ against $ln\left(\widehat{\sigma}_t^2\right)$ computed using estimates of the TVGARCH(1,1)-in-mean model. Graphs on the second row of Figure 4.14 plots the risk premium, $\widehat{\lambda}_t\widehat{\sigma}_t$, versus $ln\left(\widehat{\sigma}_t^2\right)$. The plots on the left hand side are constructed using the NL-ILS estimator computed computed with the Epanechnikov kernel function, whereas the graphs on the center and on the right hand side display scatter plots computed using the NL-ILS estimator computed with the Gaussian and the flat kernel functions, respectively.

Table 4.1: Bootstrap performance: Coverage probability and RMSD

	Parametric Bootstrap			Wild Bootstrap		
CI	Epanechnikov	Gaussian	Flat	Epanechnikov	Gaussian	Flat
90%	0.932 (0.386)	0.898 (0.384)	0.937 (0.388)	0.939 (0.388)	0.846 (0.386)	0.871 (0.384)
95%	0.960 (0.463)	0.955 (0.458)	0.989 (0.463)	0.975 (0.464)	0.912 (0.458)	0.937 (0.457)
99%	0.995 (0.616)	0.981 (0.610)	1.000 (0.618)	1.000 (0.604)	0.983 (0.598)	0.996 (0.598)

All measures are computed using the kernel based NL-ILS estimator. We generate a TVGARCH(1,1)-in-mean models as in (4.9) and (4.11), where $\eta_t \sim N(0,1)$. The time-varying parameter λ_t is set to be a bounded random walk process as in (4.5) with $c = 0.9$. We model ξ_t as a AR(1) process such that $\xi_t = \rho\xi_{t-1} + \varsigma_t$, where $\rho = 0.7$ and $\varsigma_t \sim N(0, 0.02)$. The parameters governing the conditional variance equation is set to be equal to $\phi = (0.01, 0.05, 0.9)'$. We set the number of bootstrap replications B equal to 1000 and $H = T^{0.5}$. We report the coverage probability associated with different CI's. Measures inside the brackets are the root mean squared distance (RMSD) between the upper and the lower bound defined by the confidence interval computed using the bootstrap framework.

Table 4.2: TVGARCH(1,1)-in-mean: λ_t as a bounded random walk

		λ_t		RMSE			Mean		
Bandwidth - H	Kernel	RMSE	Corr	ω	α	β	ω	α	β
$T^{0.2}$	Epanechnikov	0.39	0.60	0.01	0.03	0.10	0.02	0.07	0.84
$T^{0.2}$	Gaussian	0.25	0.71	0.01	0.02	0.06	0.01	0.05	0.88
$T^{0.2}$	Flat	0.36	0.61	0.01	0.02	0.07	0.01	0.06	0.86
$T^{0.3}$	Epanechnikov	0.26	0.72	0.01	0.02	0.05	0.01	0.05	0.88
$T^{0.3}$	Gaussian	0.18	0.81	0.01	0.02	0.05	0.01	0.05	0.89
$T^{0.3}$	Flat	0.24	0.73	0.01	0.02	0.05	0.01	0.05	0.89
$T^{0.4}$	Epanechnikov	0.18	0.80	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.4}$	Gaussian	0.13	0.86	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.4}$	Flat	0.17	0.81	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.5}$	Epanechnikov	0.13	0.86	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.5}$	Gaussian	0.11	0.88	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.5}$	Flat	0.13	0.86	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.6}$	Epanechnikov	0.12	0.88	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.6}$	Gaussian	0.13	0.86	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.6}$	Flat	0.12	0.86	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.7}$	Epanechnikov	0.13	0.85	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.7}$	Gaussian	0.16	0.79	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.7}$	Flat	0.15	0.81	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.8}$	Epanechnikov	0.17	0.77	0.01	0.02	0.05	0.01	0.05	0.89
$T^{0.8}$	Gaussian	0.21	0.68	0.01	0.02	0.05	0.01	0.05	0.89
$T^{0.8}$	Flat	0.19	0.67	0.01	0.02	0.05	0.01	0.05	0.89

All measures are computed using the kernel based NL-ILS estimator. We generate a TVGARCH(1,1)-in-mean models as in (4.9) and (4.11), where $\eta_t \sim N(0,1)$. The time-varying parameter λ_t is set to be a bounded random walk process as in (4.5) with $c = 0.9$. We model ξ_t as a AR(1) process such that $\xi_t = \rho\xi_{t-1} + \varsigma_t$, where $\rho = 0.7$ and $\varsigma_t \sim N(0, 0.02)$. The parameters governing the conditional variance equation is set to be equal to $\phi = (0.01, 0.05, 0.9)'$. MSE and RMSE account for meas squared error, root mean squared error, respectively, whereas Corr is the point-wise correlation between λ_t and $\widehat{\lambda}_t$.

Table 4.3: TVGARCH(1,1)-in-mean: λ_t as an AR(1) process

		λ_t		RMSE			Mean		
Bandwidth - H	Kernel	RMSE	Corr	ω	α	β	ω	α	β
$T^{0.2}$	Epanechnikov	0.40	0.33	0.01	0.02	0.09	0.02	0.07	0.84
$T^{0.2}$	Gaussian	0.30	0.31	0.01	0.02	0.06	0.01	0.05	0.88
$T^{0.2}$	Flat	0.38	0.31	0.01	0.02	0.07	0.01	0.06	0.86
$T^{0.3}$	Epanechnikov	0.30	0.30	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.3}$	Gaussian	0.25	0.25	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.3}$	Flat	0.29	0.25	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.4}$	Epanechnikov	0.26	0.23	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.4}$	Gaussian	0.24	0.18	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.4}$	Flat	0.26	0.17	0.01	0.02	0.05	0.01	0.04	0.90
$T^{0.5}$	Epanechnikov	0.24	0.16	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.5}$	Gaussian	0.23	0.12	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.5}$	Flat	0.24	0.11	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.6}$	Epanechnikov	0.23	0.10	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.6}$	Gaussian	0.22	0.08	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.6}$	Flat	0.23	0.07	0.01	0.02	0.06	0.01	0.05	0.89
$T^{0.7}$	Epanechnikov	0.23	0.07	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.7}$	Gaussian	0.22	0.06	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.7}$	Flat	0.23	0.05	0.01	0.02	0.05	0.01	0.05	0.90
$T^{0.8}$	Epanechnikov	0.22	0.05	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.8}$	Gaussian	0.22	0.04	0.01	0.02	0.06	0.01	0.05	0.90
$T^{0.8}$	Flat	0.22	0.03	0.01	0.02	0.06	0.01	0.05	0.90

All measures are computed using the kernel based NL-ILS estimator. We generate a TVGARCH(1,1)-in-mean models as in (4.9) and (4.11), where $\eta_t \sim N(0,1)$. The time-varying parameter λ_t is set to be covariance stationary process. We model λ_t as a AR(1) process such that $\lambda_t = \rho\lambda_{t-1} + \varsigma_t$, where $\rho = 0.9$ and $\varsigma_t \sim N(0, 0.02)$. The parameters governing the conditional variance equation is set to be equal to $\phi = (0.01, 0.05, 0.9)'$. MSE and RMSE account for meas squared error, root mean squared error, respectively, whereas Corr is the point-wise correlation between λ_t and $\widehat{\lambda}_t$.

Table 4.4: Descriptive statistics

	Mean	Median	Std. Dev.	Kurtosis	N. Obs	Start Date	End Date
CRSP	0.0010	0.0026	0.0227	9.0	2,426	05/07/1963	30/09/2011
CRSP	0.0061	0.0096	0.0545	10.4	1,023	01/07/1926	01/08/2011

Table 4.4 displays the descriptive statistics for weekly and monthly frequencies. The null hypothesis in the Jarque-Bera test is reject in all indices and frequencies.

Table 4.5: US Business Cycles

Peak month	Trough month	Duration, peak to trough	Duration, trough to peak	Duration, peak to peak	Duration, trough to trough
October 1926	November 1927	13	27	41	40
August 1929	March 1933	43	21	34	64
May 1937	June 1938	13	50	93	63
February 1945	October 1945	8	80	93	88
November 1948	October 1949	11	37	45	48
July 1953	May 1954	10	45	56	55
August 1957	April 1958	8	39	49	47
April 1960	February 1961	10	24	32	34
December 1969	November 1970	11	106	116	117
November 1973	March 1975	16	36	47	52
January 1980	July 1980	6	58	74	64
July 1981	November 1982	16	12	18	28
July 1990	March 1991	8	92	108	100
March 2001	November 2001	8	120	128	128
December 2007	June 2009	18	73	81	91

Table 4.5 displays US Business Cycle Expansions and Contractions.Contractions (recessions) start at the peak of a business cycle and end at the trough. Source: National Bureau of Economic Research (NBER), Inc. 1050 Massachusetts Avenue Cambridge MA 02138 USA.

Chapter 5

Conclusion

This thesis covers four different branches of the applied and financial econometrics literature, having the use of iterative estimators as a bridge linking these topics. As research outputs, we present contributions on both empirical and econometric theory fields. Regarding the methodological contributions, we adopt three variants of iterative estimators (the iterative least squares (IOLS), the nonlinear iterative least squares (NL-ILS) and the kernel based NL-ILS estimator) to overcome estimation issues related with vector autoregressive moving average (VARMA) and volatility models, such as GARCH, GARCH-in-mean and TVGARCH-in-mean models. We establish the consistency and the asymptotic distribution of the IOLS and NL-ILS estimators considering univariate specifications (ARMA(1,1) and GARCH(1,1)) and discuss the validity of high level assumptions required to extend the theoretical results to more complex specifications (VARMA, GARCH-in-mean and TVGARCH-in-mean models). In general lines, our empirical contributions shed light on the validity of VARMA models as powerful tools to improve forecast accuracy when dealing with large

datasets. Regarding contributions to the financial econometrics literature, we document that estimates of the risk premium parameter obtained with the NL-ILS estimator are statistically significant when the CRSP index is used. We also find evidences that the risk premium parameter is time-varying, suggesting that this variable tracks and anticipates *bear market* phases and business cycles.

More precisely, Chapter 2 addresses the issue of forecasting key macroeconomic variables using large datasets using VARMA models. We overcome the estimation problem associated with the use of maximum likelihood estimator on this class of models by adopting the IOLS estimator. We establish the consistency and the asymptotic distribution considering the univariate ARMA(1,1) model and we argue that this result can be extended to VARMA models. We present an extensive Monte Carlo study showing that the IOLS estimator is feasible and presents good performance in finite sample even when dealing with high dimensional models, such as when the number of variables is equal to twenty. Furthermore, we show that under such dimensions, the MLE estimator is no longer a feasible alternative. We present promising results, showing that VARMA models estimated with the IOLS estimator are able to outperform the benchmark competitor (the AR(1) specification) under a variety of scenarios.

Chapter 3 caries contribution on the financial econometrics field, covering theoretical and empirical aspects. Firstly, we proposes a new robust estimator for GARCH-type models: the nonlinear iterative least squares (NL-ILS). We show that the NL-ILS estimator is generic enough to accommodate a variety of volatility models generally adopted in the literature. Furthermore, we show that the NL-ILS estimator is especially useful

on specifications where errors have some degree of dependence over time. This turns to be a remarkable point on the financial econometrics literature, since we show that the NL-ILS relaxes the assumption that the disturbances associated with the volatility models are martingale difference sequence processes. We illustrate the NL-ILS estimator by providing algorithms that consider the GARCH(1,1), weak-GARCH(1,1), GARCH(1,1)-in-mean and RealGARCH(1,1)-in-mean models. We establish the consistency and asymptotic distribution of the NL-ILS estimator, in the case of the GARCH(1,1) model under assumptions that are compatible with the QMLE estimator. The consistency result is extended to the weak-GARCH(1,1) model and a further extension of the asymptotic results to the GARCH(1,1)-in-mean case is also discussed. A Monte Carlo study provides evidences that the NL-ILS estimator is consistent and outperforms the MLE benchmark in a variety of specifications. Moreover, when the conditional variance is misspecified, the MLE estimator delivers biased estimates of the parameters in the mean equation, whereas the NL-ILS estimator does not. The empirical application investigates the risk premium on the CRSP, S&P500 and S&P100 indices considering different sampling frequencies. By adopting the NL-ILS estimator, we document the risk premium parameter is statistically significant only for the CRSP index. We argue that this result comes from the wider composition of the CRPS index, resembling the market more accurately, when compared to the S&P500 and S&P100 indices. This finding holds on daily, weekly and monthly frequencies and it is corroborated by a series of robustness checks.

Finally, Chapter 4 addresses the issue of misspecification of the risk premium function. Differently from the semiparametric GARCH-in-mean

literature, we assume linearity on the relation between the conditional standard deviation and the risk premium parameter, but we allow the latter to be time-varying and evolve as a random walk process. To accommodate this feature, we introduce the time-varying GARCH-in-mean (TVGARCH-in-mean) model and show that the time-varying risk premium parameter, λ_t, can be estimated using the kernel based NL-ILS estimator. A Monte Carlo study shows that the kernel based NL-ILS estimator provides accurate estimates of λ_t. We also describe a bootstrap strategy to compute the empirical confidence intervals and we show that they present good coverage probability. Using weekly and monthly data on the excess returns computed using the CRSP index, we find evidences that the risk premium parameter is time-varying. On the monthly frequency, estimates of λ_t turn to be statistically different from zero in almost half of the observations. Pointwise analyses using weekly estimates of λ_t show that the time-varying risk premium parameter picks on periods prior to financial crises and economic downturn and gets negative when market volatility increases substantially.

Bibliography

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modelling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

Asai, M., M. Mcaleer, and M. C. Medeiros (2011): "Asymmetry and Long Memory in Volatility Modeling," *Journal of Financial Econometrics*, 10(3), 495–512, Advance Access published.

Athanasopoulos, G., D. Poskitt, and F. Vahid (2007): "Two canonical VARMA forms: Scalar component models vis--vis the Echelon form," Monash Econometrics and Business Statistics Working Papers 10/07, Monash University, Department of Econometrics and Business Statistics.

Athanasopoulos, G., and F. Vahid (2008): "VARMA versus VAR for Macroeconomic Forecasting," *Journal of Business & Economic Statistics*, 26(2), 237–252.

Baillie, R. T., and R. P. DeGennaro (1990): "Stock Returns and Volatility," *Journal of Financial and Quantitative Analysis*, 25, 203–214.

Banbura, M., D. Giannone, and L. Reichlin (2007): "Bayesian VARs with Large Panels," CEPR Discussion Papers 6326, C.E.P.R. Discussion Papers.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120(1), 387–422.

Bollerslev, T. (1986): "Generalised Autoregressive Condidional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (2008): "Glossary to ARCH (GARCH)," Discussion paper, School of Economics and Management, University of Aarhus.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992): "ARCH modeling in finance: A review of the theory and empirical evidence," *Journal of Econometrics*, 52, 5–59.

Burden, R. L., and J. D. Faires (1993): *Numerical Analysis*. PWS-Kent Pub. Co.

Camba-Mendez, G., and G. Kapetanios (2004): "Bootstrap Statistical Tests of Rank Determination for System Identification," *IEEE Transactions on Automatic Control*, 49, 238–243.

Campbell, J. Y., and L. Hentschel (1992): "No news is good news: An asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, 31, 281–318.

Carriero, A., G. Kapetanios, and M. Marcellino (2008): "Forecasting with Dynamic Models using Shrinkage-based Estimation," Working Papers 635, Queen Mary, University of London, School of Economics and Finance.

——— (2011): "Forecasting large datasets with Bayesian reduced rank multivariate models," *Journal of Applied Econometrics*, 26, 735–761.

Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2012): "Semiparametric inference in a GARCH-in-mean model," *Journal of Econometrics*, 167, 458–472.

Conrad, C., and E. Mammen (2008): "Nonparametric Regression on Latent Covariates with an Application to Semiparametric GARCH-in-Mean Models," Discussion paper, University of Heilderberg.

Dahl, C. M., and E. M. Iglesias (2009): "Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary," Discussion paper, Center for Research in Econometric Analysis of Time Series - CREATES.

Davidson, J. (1994): *Stochastic Limit Theory. An Introduction for Econometricians*. Oxford University Press.

De Mol, C., D. Giannone, and L. Reichlin (2006): "Forecasting Using a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components?," CEPR Discussion Papers 5829, C.E.P.R. Discussion Papers.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3(1), 1–100.

Dominitz, J., and R. Sherman (2005): "Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation," *Econometric Theory*, 21, 838–863.

Drost, F. C., and T. E. Nijman (1993): "Temporal Aggregation of Garch Processes," *Econometrica*, 61(4), 909–927.

Drost, F. C., and B. J. M. Werker (1996): "Closing the GARCH gap: Continuous time GARCH modeling," *Journal of Econometrics*, 74, 31–57.

Engle, R. (2002): "New Frontiers for ARCH models," *Journal of Applied Econometrics*, 17, 425–446.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987): "Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model," *Econometrica*, 55, 391–407.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalised Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82(4), 540–554.

Francq, C., and J.-M. Zakoian (2000): "Estimating Weak Garch Representations," *Econometric Theory*, 16, 692–728.

——— (2008): "A Tour in the Asymptotic Theory of GARCH Estimation," Discussion Paper 2008-03, CREST.

Francq, C., and J.-M. Zakoian (2010): *GARCH Models: Structure, Statistical Inference and Financial Appliations.* John Wiley & Sons Ltd.

French, K. R., G. Schwert, and R. F. Stambaugh (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 19, 3–29.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005): "There is a risk-return trade-off after all," *Journal of Financial Economics*, 76, 509–548.

Giraitis, L., G. Kapetanios, and T. Yates (2010): "Inference on stochastic time-varying coefficient models," Discussion paper.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48, 1779–1801.

Greene, W. H. (2008): *Econometric Analysis - Sixth Edition.* Pearson Education.

Hamilton, J. (1994): *Time Series Analysis.* Princeton University Press.

Hannan, E. J., and L. Kavalieris (1984a): "A Method for Autoregressive-Moving Average Estimation," *Biometrika*, 71(2), 273–280.

——— (1984b): "Multivariate Linear Time Series Models". Advances in Applied Probability," *Advances in Applied Probability*, 16(3), 492–561.

Hansen, P. R., Z. Huang, and H. H. Shek (2012): "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility," *Journal of Applied Econometrics*, 27, 877–906.

Judd, K. L. (1998): *Numerical Methods in Economics*. MIT Press.

Kapetanios, G. (2003): "A Note on an Iterative Least Squares Estimation Method for ARMA and VARMA models," *Economics Letters*, 79(3), 305–312.

——— (2008): "Bootstrap-based tests for deterministic time-varying coefficients in regression models," *Computational Statistics & Data Analysis*, 53, 534–45.

Kapetanios, G., V. Labhard, and S. Price (2006): "Forecasting using predictive likelihood model averaging," *Economics Letters*, 91(3), 373–379.

Lettau, M., and S. C. Ludvigson (2010): *Handbook of Financial Econometrics*, vol. 1. Elsevier B.V.

Linton, O., and B. Perron (2003): "The Shape of the Risk Premium: Evidence From a Semiparametric Generalised Autoregressive Condidional Heteroscedasticity Model," *Journal of Business & Economic Statistics*, 21, 354–367.

Linton, O., and A. Sancetta (2009): "Consistent estimation of a general nonparametric regression function in time series," *Journal of Econometrics*, 152, 70–78.

Litterman, R. B. (1986): "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business & Economic Statistics*, 4(1), pp. 25–38.

Lütkepohl, H. (2007): *New Introduction to Multivariate Time Series Analysis*. Springer- Verlag.

Mcaleer, M., and M. C. Medeiros (2008): "Realized Volatility: A Review," *Econometric Reviews*, 27, 10–45.

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41(5), 867–887.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

Ng, S., and P. Perron. (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281.

Orbe, S., E. Ferreira, and J. Rodriguez-Poo (2005): "Nonparametric estimation of time varying parameters under shape restrictions," *Journal of Econometrics*, 126, 53–77.

Pagan, A. R., and Y. S. Hong (1990): "Non-Parametric Estimation and the Risk Premium," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth Nternational Symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, pp. 51–75. Cambridge University Press.

Robinson, P. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistics Analysis and Forecasting of Economic Structural Change.*, ed. by P. Hackl, chap. 15, pp. 253–264. Springer, Berlin.

Rossi, A., and A. Timmermann (2010): "What is the Shape of the Risk-Return Relation?," Discussion paper, The Rady School of Management, University of California, San Diego.

Shepard, N., and K. Sheppard (2010): "Realising the future: forecasting with high frequency based volatility (HEAVY) models," *Journal of Applied Econometrics*, 25, 197–231.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Stock, J. H., and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167–1179.

Stock, J. H., and M. W. Watson (2005): "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research, Inc.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the

Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), pp. 267–288.

VERONESI, P. (2000): "How Does Information Quality Affect Stock Returns," *Journal of Finance*, 55, 807–837.