Essays in Long Memory

Vera-Valdés, J. Eduardo

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Vera-Valdés, J. E. (2016). *Essays in Long Memory*. Institut for Økonomi, Aarhus Universitet.

ESSAYS IN LONG MEMORY

By J. Eduardo Vera-Valdés

PhD dissertation submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of the PhD degree
in Economics and Business Economics

November 2016

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS ·

CREATES
Center for Research in Econometric Analysis of Time Series

Preface

This dissertation is the result of my PhD studies at the Department of Economics and Business Economics at Aarhus University and was written in the period from September 2013 to August 2016. I am grateful to the Department of Economics and Business Economics as well as CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78) -, funded by the Danish National Research Foundation, for providing both a unique research environment and financial support.

I would like to thank my supervisor, Prof. Niels Haldrup, for his guidance, continuous support, and his insightful contributions to all chapters of this dissertation. I could not have asked for a better supervisor, our regular meetings greatly improved the final product. I would also like to thank Daniela Osterrieder and Daniel Ventosa-Santaulària for the many useful debates that resulted in a great chapter for the dissertation. Thanks in particular to Daniel for introducing me to the topic of long memory.

From February to April of 2016, I had the privilege of visiting the Erasmus University Rotterdam. I would like to thank the Faculty and Administrative Staff at the Department for making it a productive stay. Special thanks to Michel van der Wel and Dick van Dijk for their help in making it happen and for the interesting conversations.

The PhD would not have been possible if not for the vast help of Solveig Nygaard Sørensen proofreading the dissertation and with all administrative matters. Thanks for helping me navigate through Danish waters.

Life outside the PhD would not have been as meaningful if not for my family. All my love to my parents Marcela and Javier for their enormous affection that can be felt across continents, to my uncle Fernando and my aunt Laura for providing the springboard for this project, and to my beloved wife Andrea for her unconditional support and for sharing this journey with me.

Finally, I am very grateful to my friends at CREATES and at Lærdansk for the enjoyable activities, particularly to Christian, Felipe, Laura, Orimar, Ulises and Vladimir.

J. Eduardo Vera-Valdés
Aarhus, August 2016

Updated preface

The pre-defence meeting was held on October 19, 2016 in Aarhus. I would like to thank the assessment committee consisting of Javier Hualde, Universidad Pública de Navarra; Paolo Santucci de Magistris (Chair), Aarhus University; and Michel van der Wel, Erasmus University Rotterdam; for their careful reading of the dissertation and their insightful comments and suggestions. Some of the suggestions have been incorporated into the present version of the dissertation while others remain for future research.

José Eduardo Vera Valdés
Aarhus, November 2016

CONTENTS

Summary

This dissertation comprises three self contained chapters on the analysis of long memory. Long memory deals with the study of series with autocorrelations declining at a slower pace than for $ARMA$ processes. In particular, shocks to series that show long memory tend to remain relevant for longer periods of time. Thus, the presence of long memory in the series has implications for modelling, estimation, and forecasting. This dissertation contributes to all of these branches of analysis.

In the first chapter, coauthored with Niels Haldrup, we study one of the main theoretical motivations behind the presence of long memory in time series data. Granger (1980) showed that if a series is the result of cross-sectional aggregation of persistent micro units with random coefficients, then it will show hyperbolic decaying autocorrelations. That is, the resulting series will show long memory in the covariance sense. We extend this result to other definitions of long memory considered in the time series literature. Furthermore, via Monte Carlo simulations, we examine the finite sample properties of the cross-sectional aggregation result. We find that the cross-section dimension must increase at a similar rate as the sample size for the long memory result to hold. Moreover, the degree of memory tends to be exaggerated in finite samples, particularly for low degrees of memory. By computing the autocorrelation function of a fractionally differenced cross-sectional aggregated series, we show that the long memory generated by cross-sectional aggregation does not belong to the $ARFIMA$ class of processes. Nonetheless, the fractionally differenced series has absolutely summable autocorrelation function and thus it belongs to the class of short memory processes.

The second chapter examines the performance of the $ARFIMA$ class of models when forecasting long memory series generated by sources other than $ARFIMA$ models. We use the Model Confidence Set approach of Hansen, Lunde, and Nason (2011) to compare the forecasting performance of $ARFIMA$ model specifications against $ARMA$ and high-order AR models when modelling long memory time series. As sources of memory, we consider cross-sectional aggregation of persistent micro

units (Granger, 1980), and the inclusion of shocks of random duration (Parke, 1999). We find that the $ARFIMA$ class of models is well suited for forecasting long memory at long horizons, while being competitive at shorter horizons. Moreover, we compare the forecasting performance of the heterogenous autoregressive model (HAR) of Corsi (2009) against unconstrained same-order AR models and find that the restrictions imposed by the HAR model improves forecasting performance at long horizons and for higher degrees of memory, at the cost of reduced forecasting performance at short horizons.

The third and final chapter, coauthored with Daniela Osterrieder and Daniel Ventosa-Santaulària, analyzes the estimation of an unbalanced regression product of an induced long memory corruption in the data. In Finance, the Capital Asset Pricing Model ($CAPM$) implies that financial market participants care about risk and adjust their return expectations accordingly. That is, expected returns can be explained by risk. Typically, this relation is modelled by a linear equation relating a risk measure against expected returns. Nonetheless, risk measures are found to possess long memory, while the expected returns are short memory; thus implying an unbalanced regression. In this context, we propose a Data Generating Process (DGP) that is able to capture this phenomenon. We assume that the risk series observed are corrupted by a long memory component product of breaks or cross-sectional aggregation. We show that the OLS estimate of this regression is inconsistent, but standard inference is possible. To obtain a consistent slope estimate, we propose a method that filters the long memory error component without fractional differencing. We prove that the product of a short memory process and a long memory process eliminates the long memory behavior. We then propose to use this device in an IV setting to obtain consistent estimators. Furthermore, we prove that Sargan's test for instrument validity remains valid in this unbalanced set-up. Applying the procedure to the prediction of daily returns on the S&P 500, our empirical analysis confirms return predictability and a positive risk-return trade-off.

References

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174–196.

Granger, C. W., 1980. Long memory relationships and the aggregation of dynamic models. Journal of econometrics 14, 227–238.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Parke, W., 1999. What is fractional integration? Review of Economics and Statistics 81 (4), 632–638.

DANISH SUMMARY

Denne afhandling består af tre kapitler, der fokuserer på analyse af tidsseriedata med lang hukommelse – long memory. Long memory i tidsserier kommer til udtryk ved at autokorrelationsfunktionen for processen aftager meget langsomt, (langsommere end for $ARMA$ processer) således, at observationer målt langt tilbage i tid tenderer til at være forholdsvis højt korreleret med observationer målt i dag. Sådanne egenskaber har vigtige implikationer for økonometrisk modellering, estimation og prognosefremskrivninger. Denne afhandling beskæftiger sig med alle disse emner.

I det første kapitel med Professor Niels Haldrup som medforfatter belyser vi en af de teoretiske motivationer for tilstedeværelsen af long memory. Granger (1980) har vist, at hvis en observeret tidsserieproces er fremkommet ved aggregering over mange dynamiske mikroenheder og parametrene for de bagvedliggende mikroprocesser er trukket fra en Beta-fordeling, så vil den aggregerede proces udvise long memory målt ved en hyperbolsk aftagende autokorrelationsfunktion. Vi viser, at også andre definitioner af long memory er opfyldt for at aggregeringsargumentet vil holde. Vi undersøger aggregeringsresultatets implikationer i små stikprøver målt i både cross-section og tidsserie-dimensionen, og vi dokumenterer, at et betydeligt antal observationer i begge dimensioner er nødvendigt for at kunne måle de teoretiske egenskaber. Endelig viser vi, at selvom den underliggende proces vil følge en fraktionel Brownsk bevælgelse i grænsen, så har den fraktionelle differens af processen en meget kompliceret dynamik der i særdeleshed ikke kan modelleres som en lineær $ARMA$ proces. Dette resultat har mulige implikationer for parametrisk modellering af long memory processer.

Det andet kapitel i afhandlingen ligger i forlængelse af analyserne i Kapitel 1. Her undersøges specielt, hvorledes $ARFIMA$ klassen af processer er brugbar for fremtidige prediktioner, når den bagvedliggende kilde til long memory er enten en cross-section aggregeret tidsserie proces (med long memory som diskuteret ovenfor) eller som en såkaldt "Error Duration Model" der fremkommer ved aggregering af stød, der alle har en begrænset stokastisk levetid, se Parke (1999). For en bred klasse af $ARFIMA$

processer benyttes “Model Confidence Set” til at beskrive, hvilke modelspecifikationer der bedst synes at beskrive data målt ved modellernes forecast-egenskaber. Vi finder, at $ARFIMA$ modeller specielt egner sig til langsigts-prediktioner og klarer sig nogenlunde for en kortere tidshorisont. Vi finder også, at restringerede (lange) AR modeller, såkaldte HAR modeller, specielt egner sig til forecast på langt sigt og knap så godt på kort sigt.

Det tredje og sidste kapitel har et lidt andet fokus. Kapitlet har Daniela Osterrieder og Daniel Ventosa-Santaulària som medforfattere. Vi fokuserer på estimation af såkaldte ubalancerede regressionsmodeller, når persistensen af de underliggende variable kan karakteriseres ved en kombination af long memory og short memory processer. I finansiering implicerer den såkaldte $CAPM$ model, at investorer tager højde for aktivers risiko og tilpasser deres forventninger til afkast hertil. Sammenhængen mellem et mål for risiko og forventet afkast modelleres typisk ved en lineær regression. Imidlertid finder man ofte, at risikomålet udviser long memory, imens det forventede afkast er short memory, hvilket er baggrunden for den ubalancerede regression. Vi foreslår en modelramme, der kan håndtere dette problem. Vi antager, at risikoserien er kontamineret med støj, der udviser long memory forårsaget af f.eks. strukturelle skift eller cross-section aggregering. Under disse antagelser vises OLS at føre til inkonsistente parameterskøn. Men det er stadig muligt at lave standard inferens. Vi foreslår en IV metode, der filtrerer fejlkomponentet med long memory uden at tage en fraktionel differens af serien. Vi viser, at Sargans test for instrument-validitet forbliver gyldigt i den model, vi benytter. Sluttelig anvender vi metoden til forudsigelse af daglige afkast for S&P500. Studiet bekræfter afkast-prediktabilitet og en positiv risiko-forventet afkast-trade-off.

CHAPTER 1

LONG MEMORY, FRACTIONAL INTEGRATION, AND CROSS-SECTIONAL AGGREGATION

Niels Haldrup
Aarhus University and CREATES
J. Eduardo Vera-Valdés
Aarhus University and CREATES

Abstract

It is commonly argued that observed long memory in time series and financial variables can result from cross-sectional aggregation of dynamic heterogenous micro units. In this paper, we demonstrate that the aggregation argument is consistent with a range of different long memory definitions. In a simulation study, we show however that both the cross-section and time dimensions have to be rather large to reflect the true implied memory when using commonly used estimators, especially when the theoretical memory is not too high. Finally, we show that even though the aggregated process will converge to a generalized fractional Brownian motion in the limit, the fractionally differenced series will still have an autocorrelation function that exhibits hyperbolic decay, but at a rate that still ensures summability. The fractionally differenced series is thus $I(0)$ but standard $ARFIMA$ modelling may be invalid when the long memory is caused by aggregation.

1.1 Introduction

Without specifically talking about long memory, the study of this concept in time series goes back to Granger (1966) in his article about the spectral shape near the origin for economic time series variables. He found that *long-term fluctuations, if decomposed into frequency components, are such that the amplitudes of the components decrease smoothly with decreasing period* (Granger, 1966, p. 155). This certainly applies for non-stationary $I(1)$ processes and more generally for the class of fractionally integrated processes as demonstrated by Granger and Joyeux (1980). Such processes have long lasting correlations that decay hyperbolically instead of the standard geometric decay characterizing $ARMA$ processes.

This kind of behavior, along similar findings in other scientific areas, has given rise to several definitions of long memory. In this study, following Guégan (2005), we consider five definitions of long memory.

Definition. *Let x_t be a stationary time series with autocovariance function $\gamma_x(k)$ and spectral density function $f_x(\lambda)$, and let $d \in (0, 1/2)$, then x_t has long memory*

(i) *in the **covariance sense** if $\gamma_x(k) \approx C_x k^{2d-1}$ as $k \to \infty$ with C_x a constant*

(ii) *in the **spectral sense** if $f_x(\lambda) \approx C_f \lambda^{-2d}$ as $\lambda \to 0$ with C_f a constant*

(iii) *in the **rate of the partial sum sense** if $\mathrm{Var}(\sum_t^T x_t) \approx C_v T^{1+2d}$ as $T \to \infty$ with C_v a constant*

(iv) *in the **self-similar sense** if $m^{1-2d}\mathrm{Cov}(x_t^{(m)}, x_{t+k}^{(m)}) \approx C_m k^{2d-1}$ as $k, m \to \infty$ where $x_t^{(m)} = \frac{1}{m}(x_{tm-m+1} + \cdots + x_{tm})$ with $m \in \mathbb{N}$, $m/k \to 0$, and C_m a constant*

(v) *in the **distribution sense** if $X_n(\xi) = \sigma_n^{-1} \sum_{t=1}^{\lfloor n\xi \rfloor} x_t \xrightarrow{d} B_H(\xi)$, where $\sigma_n^2 = \mathrm{E}[(\sum_{t=1}^n x_t)^2]$, $\xi \in [0, 1]$, $B_H(\xi)$ is a fractional Brownian motion, $H = d + 1/2$, and $\xrightarrow{d}$ denotes weak convergence in $D[0, 1]$, the space of real-valued functions that are continuous from the right on $[0, 1)$, and with finite limits from the left on $(0, 1]$*

where $g(x) \approx h(x)$ as $x \to x_0$ means that $g(x)/h(x)$ converges to 1 as x tends to x_0, and $\lfloor \cdot \rfloor$ denotes the integer value of its argument.

Definition (ii) is the feature considered by Granger (1966) in his study of the typical spectral shape for economic variables. The behavior of the spectrum near the origin is also used in the construction of one of the most popular estimators for long memory due to Geweke and Porter Hudak (1983) who proposed an estimation procedure based on semiparametric regression around the zero frequency.

Diebold and Inoue (2001) based their work on spurious long memory on definition (iii). They showed that structural breaks or regime switching schemes can be confused with long memory by focusing on the rate at which the variance of partial sums grows in time. Their paper demonstrates that certain stochastic processes are long memory by one definition but not necessarily by other definitions.

Definitions (iv) and (v) are largely based on the work of Mandelbrot and Van Ness (1968) for fractals. They defined the self-similarity condition and showed that the fractional Brownian motion in particular has this property.

Finally, definition (i), concerned with the behavior of the autocorrelation function for large lags, was one of the motivations behind the $ARFIMA$ model due to Adenstedt (1974), Granger and Joyeux (1980), and Hosking (1981). They extended the $ARMA$ model to account for fractional differencing. That is, for a stationary fractional process

$$(1-L)^d A(L) x_t = B(L)\epsilon_t, \tag{1.1}$$

where ϵ_t is a white noise process, $d \in (-1/2, 1/2)$, and $A(L)$, $B(L)$ are polynomials in the lag operator, L, with no common roots, all outside the unit circle. They used the standard binomial expansion to decompose $(1-L)^d$ in a series with coefficients $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$ for $j \in \mathbb{N}$. Using Stirling's approximation, it can be shown that these coefficients decay at a hyperbolic rate ($\pi_j \approx j^{d-1}$ as $j \to \infty$), which in turn translates to slowly decaying autocorrelations.

It is well known that $ARFIMA$ processes are long memory by definitions (i) through (iii), and an analogous derivation as in the proof of Theorem 1 below shows that it is also long memory in the self-similar sense, definition (iv). Moreover, a scaled partial sum of an $ARFIMA$ process converges to fractional Brownian motion, see for instance Davydov (1970) and Davidson and de Jong (2000). Thus, in the time series literature this has become the canonical construction for modelling long memory.

Even though the $ARFIMA$ model seems to be an appropriate specification to study long memory, the source underlying its dynamic features is still not clear. Physical (turbulence, see for instance Kolmogorov (1941)), as well as psychological reasons (Pearson (1902) *personal equation*), have been used to explain the presence of long memory. More recently, Parke (1999) proposed the error-duration model which relies on a decomposition of the time series into the sum of a sequence of shocks of stochastic magnitude and duration. He shows that if only a small proportion of the errors survive for large periods of time, then the resulting series shows long memory in the covariance sense, definition (i). Nonetheless, given the nature in which the data is collected, one of the main arguments often given in economics to why the data seems to have long memory features is due to cross-sectional aggregation. It is also commonplace to see arguments for cross-sectional aggregation motivating the presence of fractional long memory in real data.

Granger (1980), in line with the results of Robinson (1978) on random $AR(1)$ models, showed that cross-sectional aggregation of $AR(1)$ processes with random coefficients could produce long memory. Using a Beta distribution for the generation of cross-sectional $AR(1)$ coefficients, he showed that, as the cross-sectional dimension goes to infinity, the autocovariance function exhibits hyperbolic decay, rather than the standard geometric rate characterizing $ARMA$ processes. Thus, cross-sectional aggregation can produce long memory in the covariance sense, definition (i).

In this paper we focus on the aggregation argument leading to long memory. We address the particular specification considered by Granger because the Beta distribution is a rather flexible specification but the analysis could be extended to other aggregation schemes. We demonstrate that this aggregation scheme implies that the aggregated series is long memory using all the definitions considered in this paper. Since the aggregation result is an asymptotic property, we conduct a Monte Carlo simulation study to quantify how aggregation can lead to long memory in finite samples. The theoretical degree of memory of the aggregated series is tied to a particular parameter of the Beta distribution which affects the density mass around one. The simulations show that both the time series and the cross section dimensions have to be significant for the theoretical degree of memory to apply. Finite samples will still exhibit long memory but the estimated memory parameter can be rather large compared to its theoretical value, especially when the memory is only of moderate degree. In the third part of the paper, we focus on the extent to which the memory implied by aggregation can be removed by fractional differencing. In particular, we are interested in how $ARFIMA$ type of long memory models can be useful for practical model building. It occurs that the fractionally differenced series, using the theoretical degree of differencing, does remove the long memory of the process. The resulting series has absolutely summable autocorrelations and thus it is $I(0)$ by the definition of Davidson (2009). However, the fractionally differenced series will still have autocorrelations that decay hyperbolically, and hence will decay slower than what an $ARMA$ specification will be able to fit. This feature is most dominant when the degree of memory is moderate as opposed to being close to non-stationarity, $d \geq 0.5$. Our findings may have implications for the argument that is often given for estimating $ARFIMA$ models, namely that the observed long memory of time series can occur due to cross-sectional aggregation.

In section 2, the Granger aggregation scheme is presented and the features of the aggregated series are examined using the different long memory definitions that we consider. Section 3 presents the simulation study, and finally section 4 derives the features of fractional differencing of cross-sectionally aggregated long memory processes. The final section concludes.

1.2 Long Memory and Cross-Sectional Aggregation

Consider the random $AR(1)$ process given by:

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}, \tag{1.2}$$

where $\varepsilon_{i,t}$ is a white noise process independent of α_i with $\mathrm{E}[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2$, $\forall t \in \mathbb{Z}$ and $\alpha_i^2 \sim \mathscr{B}(\alpha; p, q)$ with $p, q > 1$ and $\mathscr{B}(\alpha; p, q)$ is the Beta distribution with density:

$$\mathscr{B}(\alpha; p, q) = \frac{1}{B(p,q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0,1), \tag{1.3}$$

where $B(\cdot,\cdot)$ is the Beta function.

Robinson (1978) proved that the process given by (1.2) admits a variance-covariance stationary solution. Furthermore, the unconditional autocorrelation function of this process shows hyperbolic decay. However, the process is not ergodic in the sense that random samples will depend on the realization of α_i.

Granger (1980) proposed[1] to consider the cross-sectional aggregation of series generated by (1.2). The cross-sectional aggregated series is defined by:

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}, \tag{1.4}$$

where $\varepsilon_{i,t}$ is a white noise process with $\mathrm{E}[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2$ $\forall i \in \{1, 2, \cdots, N\}$, $\forall t \in \mathbb{Z}$ and $\{\alpha_i\}_{i=1}^N$ are $i.i.d.$ with $\alpha_i^2 \sim \mathscr{B}(\alpha; p, q)$ with $p, q > 1$ $\forall i \in \{1, 2, \cdots, N\}$ and $\mathscr{B}(\alpha; p, q)$ as in (1.3). Furthermore, $\varepsilon_{i,t}$ is independent from α_i $\forall i \in \{1, 2, \cdots, N\}$, $\forall t \in \mathbb{Z}$.

Note that considering (1.4) instead of (1.2) solves the ergodicity violation by eliminating the dependence of the autocorrelation function on the particular realization of the autoregressive coefficient. Intuitively, note that if N is large enough, samples from (1.4) will have similar realizations of $\{\alpha_i\}_{i=1}^N$ and thus will have similar autocorrelation functions.

Granger showed that, as $N \to \infty$, the autocorrelations of x_t decay at a hyperbolic rate and hence generates long memory in the covariance sense according to definition

[1] Granger also considered the case with dependence across series and allowing for different variances across the cross-sectional units; for clarity, we will focus on the scenario under independence and equal variance.

(i) with parameter $d = 1 - q/2$. Taking $q \in (1,2)$, the long memory generated falls in the stationary range, $d \in (0,1/2)$. We will focus on this range for the rest of the analysis.

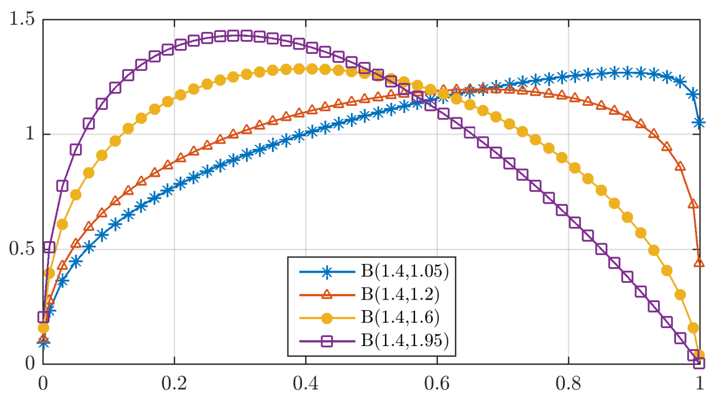
In Theorem 1, we extend the long memory result to definitions (ii) through (iv).

Theorem 1. *Let x_t be defined as in (1.4) then, as $N \to \infty$, x_t has long memory with parameter $d = 1 - q/2$ in the sense of definitions (i) through (iv).*

Proof: See appendix.

Theorem 1 shows that a cross-sectional aggregated series of infinitely many $AR(1)$ processes with squared autoregressive coefficients from a Beta distribution has long memory with long memory parameter $d = 1 - q/2$. Note that the parameters p, q are shape parameters of the Beta distribution. In particular, q affects the density around one and thus the probability of adding near unit-root $AR(1)$ processes. Furthermore, it appears that the value of p plays no role for this result as $N \to \infty$. As a consequence, Granger conjectured that asymptotically the memory only depends on the behavior of the distribution of the autoregressive coefficient near one. In Figure 1, we plot the beta distribution (1.3) for $p = 1.4$ and different values of q. As can be seen, the closer q is to one, the more density mass concentrates around one; which, as shown in Theorem 1, translates to a greater degree of memory in the cross-sectionally aggregated series, x_t.

Figure 1.1. Beta distribution.



Granger's result has been extended by, among others, Oppenheim and Viano (2004), allowing for $AR(s)$ processes (with $s \geq 1$) and Linden (1999) changing the Beta distribution to the Uniform; note that in Granger's setting the Uniform distribution was ruled out given that $p, q > 1$. Under the scenario of Oppenheim's et al., the aggregated series exhibits seasonal behavior along with long memory.

Granger's finding about the dependence of the result on the behavior of the distribution near one was further discussed by Zaffaroni (2004). He showed that if the distribution of the autoregressive coefficient, α_i, belongs to a family of absolutely continuous distributions on $[0,1)$, depending upon a real parameter $b \in (-1, \infty)$, with density

$$G(\alpha; b) \sim c_b (1-\alpha)^b \quad \text{as} \quad \alpha \to 1^-,$$

where $0 < c_b < \infty$ and 1^- denotes the limit from the left, then the aggregated series, letting $N \to \infty$, will be long memory. Moreover, the more dense the distribution of α_i is around one, the greater the degree of long memory of the aggregate. Both the Uniform and Beta distributions are members of this family of distributions. Thus, the specific parametric assumption regarding the distribution of the autoregressive coefficient is not needed for the long memory result to apply, but as we will see below, it allows us to have closed-form expressions for one of the main results in the paper. Additionally, Zaffaroni (2004) extended the result for cross-sectional aggregation to general $ARMA$ processes of finite order.

In Theorem 1, we showed that cross-sectional aggregation satisfies long memory by definitions (i) through (iv). We now argue that under one additional condition on $\varepsilon_{i,t}$, the scaled partial sum of cross-sectional aggregated series converges to fractional Brownian motion; that is, it has long memory in the distribution sense, definition (v).

$ARFIMA$ processes are fractional differenced $ARMA$ processes after adopting the $(1-L)^d$ filter. The MA series resulting from expansion of the $(1-L)^d$ filter has hyperbolically decaying coefficients of the form $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$ for $j \in \mathbb{N}$ and this produces a series with hyperbolic decaying autocovariances. We can generalize this construction to series that still show hyperbolic decaying coefficients, yet, the coefficients do not come from the fractional difference operator as defined above. We call these processes generalized fractional processes (see Davidson and de Jong (2000)).

We prove in Lemma 1 that if $\varepsilon_{i,t}$ are $i.i.d.$, cross-sectional aggregated processes can be expressed as a generalized fractional process.

Lemma 1. *Let x_t be defined as in (1.4) and assume that $\varepsilon_{i,t}$ is an $i.i.d.$ process, then, as $N \to \infty$, x_t can be expressed as*

$$x_t = \sum_{j=0}^{\infty} \phi_j \nu_{t-j},$$

where $\nu_j \sim N(0, \sigma_\varepsilon^2)$ are independent and $\phi_j = \left(B(p+j,q)/B(p,q)\right)^{1/2}$, $\forall j \in \mathbb{N}$.

Proof: See appendix.

Lemma 1 relies on the fact that, when N goes to infinity, the Central Limit Theorem can be applied. In this sense, it is in line with the work of Davidson and Sibbertsen (2005) who show that cross-sectional aggregated non-linear processes of appropriate form have linear representations in the sense of having $MA(\infty)$ representations. Note also that in Lemma 1 we could obtain a similar result if $\varepsilon_{i,t}$ is not $i.i.d.$ but satisfies Lyapunov's condition. Furthermore, the resulting series inherits the uncorrelated property of $\varepsilon_{i,t}$ and, given normality, they are independent.

By Stirling's approximation, we can show that the coefficients in the representation decay at a hyperbolic rate, $\phi_j \approx j^{-q/2} = j^{d-1}$ as $j \to \infty$ with $d = 1 - q/2$, but without being associated with the fractional differencing parameters, π_j, defined above. Thus, cross-sectional aggregated processes are generalized fractional processes. In Section 1.4, we will detail the study of the relationship between cross-sectional aggregated long memory processes and $ARFIMA$ processes.

Theorem 2 argues that the scaled partial sum of cross-sectional aggregated processes converges to fractional Brownian motion.

Theorem 2. *Let x_t be defined as in (1.4) and assume that $\varepsilon_{i,t}$ is an $i.i.d.$ process, then, as $N \to \infty$, x_t has long memory in the sense of definition (v) with parameter $d = 1 - q/2$.*

Proof: See appendix.

Theorem 2 is in line with the results from Zaffaroni (2004) when restricting the analysis to the Beta distribution. In this context, the parametric assumption allows us to find closed-form solutions for the variance terms. This in turn translates into closed-form expressions for the coefficients of the generalized fractional process. Given this, note that Theorem 2 follows directly from the developments of Davydov (1970) and Davidson and de Jong (2000).

In summary, Theorems 1 and 2 show that a cross-sectional aggregated series has long memory by all the definitions considered. However, although the coefficients of the MA representation decay hyperbolically, they are different from those arising from inversion of a fractional difference filter.

1.3 Finite Sample Study

In order to analyze the finite sample properties of Granger's aggregation result, which holds asymptotically, we conducted a Monte Carlo simulation experiment. Note that if we do not consider enough $AR(1)$ processes in the cross-sectional dimension, the resulting series may not have long memory as predicted theoretically. Granger (1990) proposed a division between cross-sectional aggregation in small scale, involving sums of a few time series variables, and large scale, involving the sums of very many variables. In particular, Chambers (1998) shows that when the number of variables is not large, the aggregation result can not be obtained. Nonetheless, the numerical finite sample implications of these conclusions should be quantified.

To shed some light on this question we generate x_t as in (1.4) under different parametric settings focusing on three main dimensions: the density of the autoregressive coefficient near one determined by the parameter q; the sample size T; and the cross-sectional dimension N, that is, the number of $AR(1)$ processes aggregated over.

The simulation proceeds as follows for R replications:

- Sample the N autoregressive coefficients from the density function, equation (1.3).
- Generate the individual $AR(1)$ series of size T, equation (1.2), using the sampled coefficients. The error terms, $\varepsilon_{i,t}$, are sampled from independent standard normals.
- Aggregate the individual series cross-sectionally according to equation (1.4).
- Estimate the long memory parameter by the GPH estimator, see Geweke and Porter Hudak (1983). For robustness, we also consider the local Whittle estimator of Robinson (1995) and Künsch (1986) [LW], and the bias-reduction method for the GPH estimator suggested by Andrews and Guggenberger (2003) using second degree [$AND(2)$] and fourth degree polynomials [$AND(4)$].

We use these estimators of the long memory parameter since they do not depend on a full parametric assumption. The importance of this will be made clearer in Section 1.4 when discussing the relationship of cross-sectional aggregated series with $ARFIMA$ processes.

Throughout, we have used a bandwidth of $T^{0.5}$ for all estimators as it is standard in the literature. As it is well known, the bandwidth affects the bias-precision trade-off. Results with different bandwidths are available upon request showing this trade-off; notwithstanding, the main conclusions maintain. Moreover, for reasons of space, we present simulations for $p = 1.4$ throughout, so that the density for the autoregressive coefficient takes the form shown in Figure 1.1. For robustness we have tried different values of p, available upon request, with similar qualitative results despite minor quantitative differences.

To analyze the importance of the density around one on the aggregation result, we report in Table 1.1 the results from the simulations for different values of q in (1.3), which is related to the degree of long memory $d = 1 - q/2$. We have conducted

$R = 10,000$ replications with $T = N = 10,000$. Additionally, for comparison, we also simulate 10,000 $FI(d)$ series using the exact algorithm of Jensen and Nielsen (2014).

Table 1.1. Mean and standard deviation in parentheses of the estimated long memory parameter. $T = N = R = 10,000$. The last three columns show comparable $FI(d)$ processes simulated according to Jensen and Nielsen (2014) algorithm.

Theoretical	Cross-sectional aggregated				$FI(d)$			
d	GPH	LW	$AND(2)$	$AND(4)$	GPH	LW	$AND(2)$	$AND(4)$
0.475	0.5117	0.5067	0.4967	0.4920	0.4818	0.4779	0.4840	0.4849
	(0.0711)	(0.0563)	(0.1126)	(0.1463)	(0.0710)	(0.0565)	(0.1116)	(0.1467)
0.45	0.4894	0.4840	0.4731	0.4671	0.4566	0.4519	0.4582	0.4606
	(0.0718)	(0.0544)	(0.1139)	(0.1499)	(0.0700)	(0.0550)	(0.1128)	(0.1471)
0.4	0.4442	0.4409	0.4255	0.4186	0.4029	0.4034	0.4045	0.4051
	(0.0723)	(0.0598)	(0.1135)	(0.1482)	(0.0699)	(0.0563)	(0.1120)	(0.1465)
0.35	0.4041	0.4031	0.3826	0.3744	0.3536	0.3504	0.3542	0.3541
	(0.0722)	(0.0578)	(0.1127)	(0.1482)	(0.0698)	(0.0572)	(0.1104)	(0.1449)
0.3	0.3633	0.3601	0.3394	0.3295	0.3017	0.3012	0.3040	0.3043
	(0.0723)	(0.0564)	(0.1155)	(0.1508)	(0.0693)	(0.0529)	(0.1102)	(0.1453)
0.25	0.3251	0.3254	0.2965	0.2829	0.2529	0.2480	0.2532	0.2529
	(0.0730)	(0.0619)	(0.1159)	(0.1520)	(0.0702)	(0.0531)	(0.1104)	(0.1442)
0.2	0.2887	0.2882	0.2573	0.2434	0.2009	0.1946	0.2012	0.2013
	(0.0738)	(0.0613)	(0.1183)	(0.1552)	(0.0700)	(0.0566)	(0.1112)	(0.1464)
0.15	0.2547	0.2517	0.2198	0.2075	0.1512	0.1458	0.1509	0.1519
	(0.0730)	(0.0619)	(0.1173)	(0.1529)	(0.0694)	(0.0537)	(0.1107)	(0.1454)
0.10	0.2252	0.2253	0.1888	0.1753	0.1004	0.0957	0.1022	0.1029
	(0.0741)	(0.0615)	(0.1174)	(0.1536)	(0.0683)	(0.0561)	(0.1103)	(0.1448)
0.05	0.1938	0.1953	0.1569	0.1422	0.0500	0.0457	0.0494	0.0493
	(0.0748)	(0.0611)	(0.1181)	(0.1550)	(0.0692)	(0.0550)	(0.1104)	(0.1472)

Note. The estimators considered are GPH, Geweke and Porter Hudak (1983), LW, the local Whittle estimator of Robinson (1995) and Künsch (1986), $AND(2)$ and $AND(4)$ are the bias corrected GPH tests of Andrews and Guggenberger (2003) using second degree and fourth degree polynomials, respectively.

The table shows that for large degrees of memory the estimates are close to the theoretical values but rather distant when the memory is low. Moreover, the estimates are rather robust to the estimation procedure. Thus, it shows that the density of the autoregressive coefficient plays a key role in finite samples. It suggests that using cross-sectional aggregation as a way to simulate long memory works poorly when dealing with a small memory index, d. In contrast, Table 1.1 shows that fractional differencing remains precise for all values of d. In particular, note that for a sample size of 10,000 and using 10,000 $AR(1)$ series, the cross-sectional aggregated series tends to show a larger degree of memory than the asymptotic result implies, and that of a comparable $FI(d)$ process.[2] This, coupled with the computational load required to generate the aggregated series, suggests that the aggregation scheme is dominated by exact fractional differencing.

[2]We need a sample size T and cross-sectional dimension N of more than 100,000 to obtain results that mimic the $FI(d)$ simulations.

Moving on to analyze the importance of the cross-sectional dimension, we present in Figure 1.2 box-plots from simulations with a sample size of $T = 10,000$ while varying the cross-sectional dimension N. For ease of exposition, we only present results for four theoretical degrees of long memory with the GPH estimation method.

Figure 1.2. Box-plot of the GPH long memory estimator for different levels of aggregation. $T = R = 10,000$. In each box the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the 95% coverage assuming symmetry.

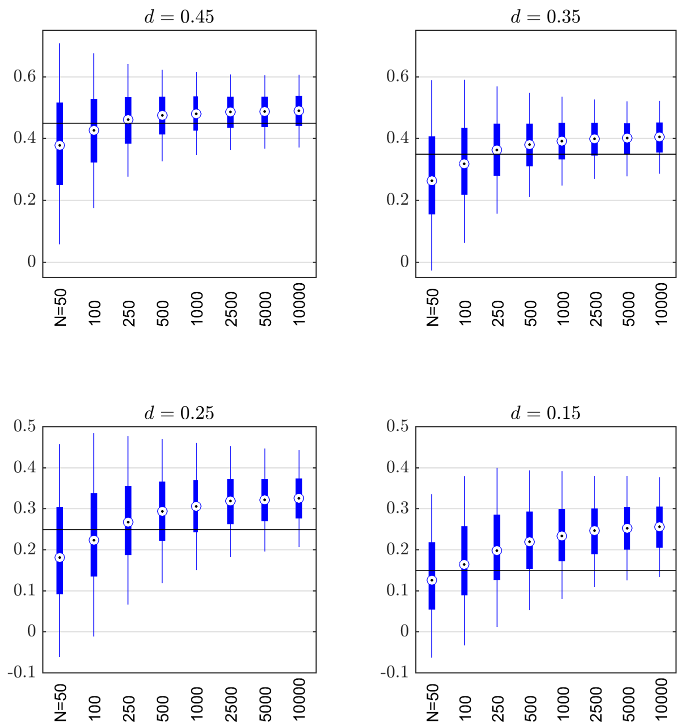
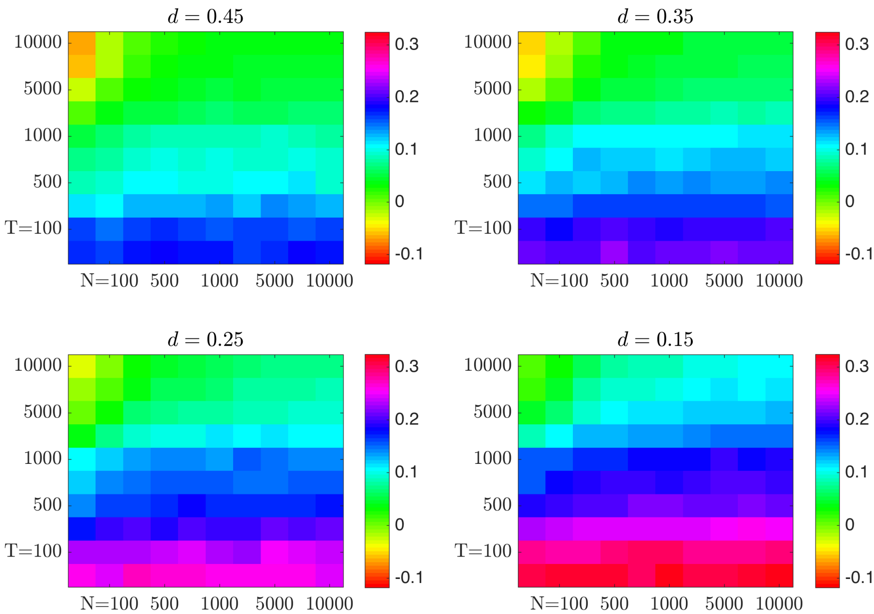


Figure 1.2 allows us to see how the long memory parameter evolves while increasing the cross-sectional dimension. It further shows the dependence of the result on the density of the autoregressive coefficient and the implied theoretical memory d. The larger the degree of memory (the denser the Beta distribution around one) the better we can approximate the asymptotic result. For small values of N, the figures show that the median is below the theoretical value in all cases, which is line with the result by Chambers (1998) on small scale aggregation. It can also be seen that the memory parameter is generally imprecisely estimated when N is relatively small. Moreover, the

box-plots show that the cutoff between small and large scale aggregation varies with the density of the autoregressive coefficients. In general, with a sample size of 10,000, for larger degrees of memory, we need at least 250 $AR(1)$ series so that the median of the simulations is close to the theoretical values, while for smaller degrees of memory, as Table 1.1 showed, we are still far away even with 10,000 $AR(1)$ series. Moreover, much estimation uncertainty is still present in all cases.

Finally, to study the interaction between the sample size and the cross-section dimension, Figure 1.3 presents the heat-maps of the mean of the GPH estimated parameters for 1,000 replications minus their theoretical values while varying T and N. We consider four theoretical values of $d \in \{0.45, 0.35, 0.25, 0.15\}$.

Figure 1.3. Heat-map of the mean of the GPH estimator for $R = 1000$ replications minus the theoretical value; $T, N \in \{50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000\}$.



The figure shows the interaction between the cross-sectional dimension and the sample size. For smaller sample sizes, we are always overshooting the true long memory parameter. This suggests that when working with a small sample size, the estimators do not have enough information to discern the true nature of the process. On the other hand, as the sample size T increases, more cross-sectional units are needed to approximate the asymptotic result. Thus, it quantifies the cutoff between small and large scale aggregation. This indicates that if we were to use aggregation as a way to simulate long memory we need to increase the cross-sectional dimension proportionally to the sample size, with the associated computational cost that it implies.

In summary, we find that the aggregation scheme to generate long memory can be rather imprecise and generally requires many time series observations and many cross-sectional units. In particular for small values of d.

1.4 Cross-Sectional Aggregation and *ARFIMA* processes

Theorems 1 and 2 together with Lemma 1 show that cross-sectional aggregated processes share key properties with $ARFIMA$ processes. Both processes satisfy all of the definitions of long memory considered in this paper, and both have $MA(\infty)$ representations with hyperbolic decaying coefficients.

These shared properties may explain why several authors have assumed that cross-sectional aggregated processes are of the $ARFIMA$ type. For instance, Balcilar (2004) and Gadea and Mayoral (2006) refer to cross-sectional aggregation as a possible explanation behind long memory found in inflation and fit $ARFIMA$ models using parametric methods.

Granger (1980), in his original article, also noted that although aggregated series were not $ARFIMA$, the $ARFIMA$ specification could provide a good approximation.

Others have suggested that the long memory of the cross-sectional aggregated series can be eliminated by fractional differencing. Diebold and Rudebusch (1989) allude to aggregation as the origin of long memory in output. They estimate the long memory parameter by the GPH method, fractionally difference the series, and subsequently estimate an $ARMA$ model. Kumar and Okimoto (2007), refer aggregation as the motive behind long memory and use the Shimotsu and Phillips (2005) estimator for the long memory parameter. This method relies on fractional differencing.

Recall from (1.1) that an $ARFIMA$ process is a fractionally differenced $ARMA$ process. Thus, if we were to take a d-th difference, $(1-L)^d$, of an $ARFIMA(a,d,b)$ process, we would recover the underlying $ARMA(a,b)$ process. However, as Lemma 1 shows, the cross-sectional aggregated process is a generalized fractional process. Thus, it may not appear from fractional differencing. As a way to give an answer to this question, Theorem 3 presents the autocovariance function of a fractionally differenced cross-sectionally aggregated process.

Theorem 3. *Let* $y_t = (1-L)^d x_t$ *where* x_t *is defined as in (1.4) with* $N \to \infty$ *and* $\gamma_y(k) =$

$E[y_t y_{t-k}]\ \forall k \in \mathbb{N}$. *Then,*

$$\gamma_y(k) = \frac{\gamma^*(k)}{B(p,q)}\left[B(p,q-1)\left(F_1(k)-1\right)+B(p+\frac{1}{2},q-1)F_2(k)\right],$$

where

$$\gamma^*(k) = \sigma_\varepsilon^2 \frac{\Gamma(1+2d)}{\Gamma(-d)\Gamma(1+d)}\frac{\Gamma(-d-k)}{\Gamma(1+d-k)},$$

is the autocovariance function of an $I(-d)$ process with innovations with variance σ_ε^2 and

$$\begin{aligned}
F_1(k) \quad &:= \quad F\left[\left\{1,p,\frac{1-d+k}{2},\frac{-d+k}{2}\right\},\left\{p+q-1,\frac{2+d+k}{2},\frac{1+d+k}{2}\right\},1\right]+\\
&\quad\quad F\left[\left\{1,p,\frac{1-d-k}{2},\frac{-d-k}{2}\right\},\left\{p+q-1,\frac{2+d-k}{2},\frac{1+d-k}{2}\right\},1\right],\\
F_2(k) \quad &:= \quad \frac{-d+k}{1+d+k}*\\
&\quad\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d+k}{2},\frac{2-d+k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d+k}{2},\frac{3+d+k}{2}\right\},1\right]\\
&\quad\quad +\frac{-d-k}{1+d-k}*\\
&\quad\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d-k}{2},\frac{2-d-k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d-k}{2},\frac{3+d-k}{2}\right\},1\right],
\end{aligned}$$

where $F[\cdot]$ is the generalized hypergeometric function.

Proof: See appendix.

Two main points can be drawn from Theorem 3.

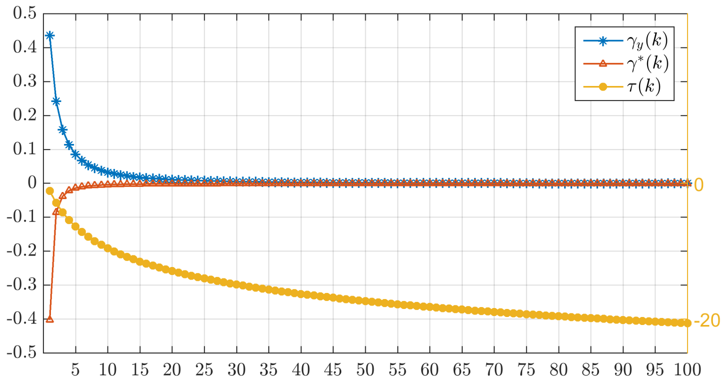
First, looking at the resulting autocovariance function, we find that it retains some memory even for large lags. In particular, it does not belong to the class of autocovariance functions for linear $ARMA$ processes. This has implications for modelling and estimation. In particular, Maximum Likelihood estimators rely on the fact that the resulting series after differencing is of the $ARMA$ type. The properties of the Quasi-Maximum Likelihood estimation of $ARFIMA$ models when the underlying process is a generalized fractional process remain an open question.

Second, note that as the proof of Theorem 3 shows, in reality we are calculating the autocovariance function of cross-sectionally aggregated $ARFIMA(1,-d,0)$ series.

Hence, the individual series are antipersistent with parameter $-d$, and the cross-sectionally aggregated AR processes are overdifferenced. The autocovariance function of the overdifferencing filter $(1-L)^d$ is given by $\gamma^*(k)$ in Theorem 3, which is a negative function in k.

Figure 1.4 displays the shape of the autocovariance function for the fractionally differenced cross-sectionally aggregated process $\gamma_y(k)$, the autocovariance of the antipersistent component $\gamma^*(k)$, and its ratio $\tau(k) := \gamma_y(k)/\gamma^*(k)$.

Figure 1.4. Autocovariance function for the fractionally differenced cross-sectionally aggregated series $\gamma_y(k)$, the $I(-d)$ process $\gamma^*(k)$ (left scale), and its ratio $\tau(k)$ (right scale). $p = 1.4$, $q = 1.05$ so that $d = 0.475$.



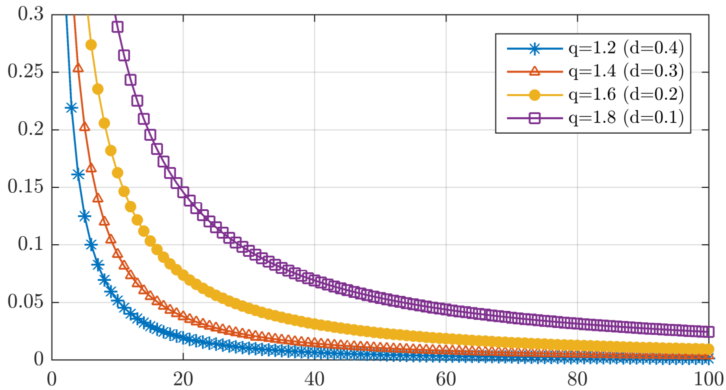
The following Corollary shows that the function $\tau(k)$ is a negative slowly varying function in k and thus the autocovariance of the fractionally differenced cross-sectionally aggregated process shows hyperbolic decay.

Corollary 1. *As $k \to \infty$, $\gamma_y(k) \approx \tau(k)k^{-1-2d}$, where $\tau(k)$ is a slowly-varying function in the sense that, for $c > 0$, $\lim_{k\to\infty} \tau(ck)/\tau(k) = 1$. Moreover, the autocorrelations are absolutely summable, that is, $\sum_{i=0}^{\infty} |\rho_y(k)| = \sum_{i=0}^{\infty} |\gamma_y(k)/\gamma_y(0)| < \infty$.*

Proof: See appendix.

As seen in Figure 1.4 and proved in Corollary 1, the autocovariance function $\gamma_y(k)$ decays at a hyperbolic rate similar to the rate for antipersistent processes. However, the sign of the function is positive as opposed to antipersistent processes, which is a feature induced by the cross-sectional aggregation. Despite the hyperbolic rate, the decay is still fast in the sense that the autocorrelations are summable and hence satisfy the condition for $I(0)$ considered by Davidson (2009).

Figure 1.5. Autocovariance functions for the fractionally differenced cross-sectionally aggregated series $\gamma_y(k)$ for $p = 1.1$ and $q \in \{1.2, 1.4, 1.6, 1.8\}$.



From the expression of $\gamma_y(k)$ given in Theorem 3, note that autocovariances for finite k depend on the parameters p and q associated with the Beta distribution. Figure 1.5 displays the autocovariance functions for $p = 1.4$ and $q \in \{1.2, 1.4, 1.6, 1.8\}$. Small values of q (and hence large memory) result in relatively small autocovariances for finite k. As q increases, and hence memory declines, the fractionally differenced series tend to have rather significant autocovariances for small as well as for moderately large lags.[3] This will clearly have a major impact on the properties of estimated parametric models of the $ARFIMA$ type which in general will be misspecified.

1.5 Conclusions

In many empirical studies, long memory is modelled as $ARFIMA$ processes and often the motivation used in this research relies on the Granger (1980) argument that cross-sectional aggregation can lead to long memory. In this paper, we argue that both $ARFIMA$ processes and long memory processes generated according to Granger's aggregation scheme satisfy a range of long memory definitions. Despite these similarities, the two classes of processes have features that are somewhat different. First of all, one should be aware that cross-sectional aggregation leading to long memory is an asymptotic feature that applies for both the cross-sectional and the time dimensions tending to infinity. In finite samples, and for moderate cross-sectional dimensions, the observed memory of the series can be rather different from the theoretical memory.

[3] We also constructed graphs similar to Figure 1.5 while varying p. They show that the autocovariances increase in size as p increases.

Moreover, the aggregation result seems to be most apparent when the memory tends to be relatively high, and hence the Beta distribution has concentrated mass around one. Secondly, we have shown that when taking a fractional difference of a cross-sectionally aggregated long memory process, the resulting process is not an $ARMA$ process. The fractionally differenced process has autocorrelations that are summable and the process is $I(0)$ according to Davidson's (2009) definition, but the autocorrelations still decay at a hyperbolic rate rather than a geometric rate. Especially when the memory is moderate the autocorrelations are more persistent than observed in $ARMA$ processes. Granger (1980) noted that cross-sectional aggregated long memory processes are likely to be well approximated as $ARFIMA$ processes in most cases. Our study shows that care should be taken regarding this common belief. In many cases, $ARFIMA$ specifications will not provide a satisfactory description of the short run dynamics even though the long memory can be effectively removed by fractional differencing.

Acknowledgments

We would like to thank the helpful comments from James Davidson and Peter Robinson, and the participants at the 4th Long Memory Symposium and the 2016 IAAE conference. The authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1.6 References

Adenstedt, R. K., 1974. On large-sample estimation for the mean of a stationary random sequence. The Annals of Statistics 2 (6), 1095–1107.

Andrews, D. W. K., Guggenberger, P., 2003. A bias–reduced log–periodogram regression estimator for the long–memory parameter. Econometrica 71 (2), 675–712.

Balcilar, M., 2004. Persistence in inflation: does aggregation cause long memory? Emerging Markets Finance and Trade 40 (5), 25–56.

Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long Memory Processes. Springer Verlag.

Chambers, M. J., 1998. Long memory and aggregation in macroeconomic time series. International Economic Review 39 (4), 1053–1072.

Davidson, J., 2009. When is a time-series I(0)? In: Castle, J. L., Shephard, N. (Eds.), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford University Press Oxford, pp. 322–342.

Davidson, J., de Jong, R. M., 2000. The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes. Econometric Theory 16 (5), 643–666.

Davidson, J., Sibbertsen, P., 2005. Generating schemes for long memory processes: regimes, aggregation and linearity. Journal of Econometrics 128 (2), 253–282.

Davydov, Y., 1970. The Invariance Principle for Stationary Processes. Theory of Probability and its applications 15 (3), 487–499.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105 (1), 131–159.

Diebold, F. X., Rudebusch, G. D., 1989. Long memory and persistence in aggregate output. Journal of Monetary Economics 24 (2), 189–209.

Gadea, M. D., Mayoral, L., 2006. The Persistence of Inflation in OECD Countries: A Fractionally Integrated Approach. International Journal of Central Banking 2, 1–55.

Geweke, J., Porter Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4 (4), 221–238.

Granger, C. W., 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14 (2), 227–238.

Granger, C. W., 1990. Aggregation of time series variables: a survey. In: Barker, T., Pesaran, M. (Eds.), Disaggregation in econometric modelling. Routledge, London, pp. 17–34.

Granger, C. W., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1 (1), 15–29.

Granger, C. W. J., 1966. The Typical Spectral Shape of an Economic Variable. Econometrica 34 (1), 150–161.

Guégan, D., 2005. How can we define the concept of long memory? An econometric survey. Econometric Reviews 24 (2), 113–149.

Hosking, J. R. M., 1981. Fractional differencing. Biometrika 68 (1), 165–176.

Jensen, A. N., Nielsen, M. Ø., 2014. A fast fractional difference algorithm. Journal of Time Series Analysis 35 (5), 428–436.

Kolmogorov, A. N., 1941. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. Doklady Akademii Nauk Sssr 30 (4), 299–303.

Kumar, M. S., Okimoto, T., 2007. Dynamics of persistence in international inflation rates. Journal of Money, Credit, and Banking 39 (6), 1457–1479.

Künsch, H., 1986. Discrimination between Monotonic Trends and Long-Range Dependence. Journal of Applied Probability 23 (4), 1025–1030.

Linden, M., 1999. Time series properties of aggregated AR (1) processes with uniformly distributed coefficients. Economics Letters 64 (1), 31–36.

Mandelbrot, B. B., Van Ness, J. W., 1968. Fractional Brownian Motions, Fractional Noises and Applications. SIAM Review 10 (4), 422–437.

Oppenheim, G., Viano, M. C., 2004. Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results. Journal of Time Series Analysis 25 (3), 335–350.

Parke, W., 1999. What is fractional integration? The Review of Economics and Statistics 81 (4), 632–638.

Pearson, K., 1902. On the mathematical theory of errors of judgment, with special reference to the personal equation. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character 198, 235–299.

Robinson, P. M., 1978. Statistical inference for a random coefficient autoregressive model. Scandinavian Journal of Statistics 5 (3), 163–168.

Robinson, P. M., 1995. Gaussian semiparametric estimation of long range dependence. The Annals of statistics 23 (5), 1630–1661.

Shimotsu, K., Phillips, P. C. B., 2005. Exact local Whittle estimation of fractional integration. The Annals of statistics 33 (4), 1890–1933.

Sowell, F., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165–188.

Zaffaroni, P., 2004. Contemporaneous aggregation of linear dynamic models in large economies. Journal of Econometrics 120 (1), 75–102.

1.7 Appendix

Proof of Theorem 1

Let x_t be defined as in (1.4).

To prove (i), note that x_t has zero mean and thus its variance is given by

$$\begin{aligned}\gamma_x(0) = E[x_t^2] &= E\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_{i,t}\right)^2\right] = \frac{1}{N}E\left[\left(\sum_{i=1}^{N} x_{i,t}\right)^2\right] \\ &= \frac{\sigma_\varepsilon^2}{N}\sum_{i=1}^{N} E\left[\frac{1}{1-\alpha_i^2}\right],\end{aligned}$$

where the third equality follows from the independence assumption.

Note that $\forall i \in \{1,2,\cdots,N\}$, unconditionally,

$$E\left[\frac{1}{1-\alpha_i^2}\right] = \int_0^1 \frac{1}{1-x}\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}dx = \int_0^1 \frac{x^{p-1}(1-x)^{q-2}}{B(p,q)}dx = \frac{B(p,q-1)}{B(p,q)},$$

which shows that each series has long memory in the covariance sense. Yet, as previously discussed, (1.2) is not ergodic in the sense that realizations depend on the draw of α_i. To solve the ergodicity violation we consider the cross-sectional aggregated series noting that,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left[\frac{1}{1-\alpha_i^2}\middle|\alpha_i\right] = \int_{-\infty}^{\infty}\frac{1}{1-\alpha_i^2}d\alpha_i = E\left[\frac{1}{1-\alpha_i^2}\right],$$

so that $\gamma_x(0) = \sigma_\varepsilon^2 B(p,q-1)/B(p,q)$ regardless of the conditioning on the autoregressive coefficients.

As for the autocovariances, similar calculations show that

$$\gamma_x(k) = E[x_t x_{t-k}] = \frac{\sigma_\varepsilon^2}{N}\sum_{i=1}^{N} E\left[\frac{\alpha_i^k}{1-\alpha_i^2}\right] = \sigma_\varepsilon^2\frac{B(p+k/2,q-1)}{B(p,q)},$$

for $k \in \mathbb{N}$, which, by Stirling's approximation,

$$\gamma_x(k) = \sigma_\varepsilon^2\frac{B(p+k/2,q-1)}{B(p,q)} = \sigma_\varepsilon^2\frac{\Gamma(q-1)}{B(p,q)}\frac{\Gamma(p+k/2)}{\Gamma(p+q+k/2-1)} \approx \sigma_\varepsilon^2\frac{\Gamma(q-1)}{B(p,q)}k^{1-q}.$$

So that the aggregated series shows hyperbolic decaying autocovariances $\gamma_x(k) \approx C_x k^{1-q}$. That is, long memory in the covariance sense with parameter $d = 1-q/2$.

To prove (ii), note that given the autocorrelation function $\rho_x(k) = \gamma_x(k)/\gamma_x(0)$ with $\gamma_x(k), \gamma_x(0)$ computed above, Theorem 1.3 in Beran, Feng, Ghosh, and Kulik (2013) shows that the spectral density has a pole in the origin.

To prove (iii),

$$\begin{aligned}
Var\left(\sum_{t=1}^{T} x_t\right) &= \frac{1}{N} E[(x_1 + x_2 + \cdots + x_T)^2] \\
&= E[x_1^2 + \cdots + x_T^2 + 2(x_1x_2 + \cdots + x_{T-1}x_T)] \\
&= TE[x_1^2] + 2E\left[\left(\sum_{t=2}^{T} x_1x_t + \cdots + \sum_{t=T-1}^{T} x_1x_t\right)\right] \\
&= TE[x_1^2] + 2\left((T-1)E[x_1x_2] + \cdots + E[x_1x_T]\right) \\
&= 2\left(\frac{T}{2} + 2\left((T-1)\gamma_x(1) + \cdots + \gamma_x(T-1)]\right)\right) \\
&\approx 2C_x\left((T-1) + (T-2)2^{1-q} + \cdots + (T-1)^{1-q}\right) \\
&= 2C_x \sum_{t=1}^{T} (T-t)t^{1-q} \approx T^{3-q} = T^{1+2d},
\end{aligned}$$

where in the previous to last line we have used the asymptotic behavior calculated in (i).

Finally, to prove (iv), we need to analyze the series while considering temporal aggregation. Let $m \in \mathbb{N}$ and define

$$x_i^{(m)} = \frac{1}{m}(x_{im-m+1} + \cdots + x_{im}),$$

for $i = \{1, 2, \cdots\}$. That is, let $x_i^{(m)}$ be a temporal aggregation of x_t at level m. Then, note that $\forall\, t \in \mathbb{N}$ and for large $k \in \mathbb{N}$

$$\begin{aligned}
E[x_t^{(m)} x_{t+k}^{(m)}] &= \frac{1}{m^2} E[(x_{tm-m+1} + \cdots + x_{tm})(x_{(t+k)m-m+1} + \cdots + x_{(t+k)m})] \\
&= \frac{1}{m^2} E[\underbrace{x_{tm-m+1}x_{(t+k)m-m+1} + \cdots + x_{tm-m+1}x_{(t+k)m} + \cdots + x_{tm}x_{(t+k)m}}_{m^2 \text{ terms}}] \\
&= \frac{1}{m^2}\left(\underbrace{\gamma_x(km) + \cdots + \gamma_x(km+m-1) + \cdots \gamma_x(km)}_{m^2 \text{ terms}}\right).
\end{aligned}$$

Factorizing terms and replacing $\gamma_x(|j-i|)$ for its asymptotic behavior calculated in (i),

$$E[x_t^{(m)} x_{t+k}^{(m)}] = \frac{1}{m^2}\left(\gamma_x(km-m+1) + \cdots + m\gamma_x(km) + \cdots + \gamma_x(km+m-1)\right)$$

$$\approx \frac{C_x}{m^2}\left((km-m+1)^{1-q}+m(km)^{1-q}+\cdots+(km+m-1)^{1-q}\right),$$

dividing both sides by k^{1-q},

$$\begin{aligned}\frac{1}{k^{1-q}}E[x_t^{(m)}x_{t+k}^{(m)}] &\approx \frac{C_x}{m^2}\left(m^{1-q}+\cdots+mm^{1-q}+\cdots+m^{1-q}\right)\\ &= \frac{C_x}{m^2}\left(1+\cdots+m+\cdots+1\right)m^{1-q}\\ &= \frac{C_x}{m^2}m^2m^{1-q}=C_xm^{1-q},\end{aligned}$$

where in the first line we used that $m/k \to 0$ as $k \to \infty$.

Thus, with $d = 1-q/2$, $m^{1-2d}\text{Cov}(x_t^{(m)}, x_{t+k}^{(m)}) \approx Ck^{2d-1}$ as $k, m \to \infty$, $m/k \to 0$.

Proofs of Lemma 1 and Theorem 2

Let x_t be defined as in (1.4). Using the infinite series representation of each $AR(1)$ process defined as in (1.2) note that x_t can be written as

$$x_t = \sum_{j=0}^{\infty}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i^j\varepsilon_{i,t-j}\right).$$

Given the additional assumption on $\varepsilon_{i,t-j}$, the classical Central Limit Theorem holds sideways and thus, $\forall j \in \mathbb{N}$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\alpha_i^j\varepsilon_{i,t-j} \sim \mathbb{N}(0,\sigma_\varepsilon^2 B(p+j,q)/B(p,q)),$$

We have used analogous derivations as in the proof above to obtain the variance terms. Note in particular that, in contrast to the proofs of Zaffaroni (2004), the parametric assumption on the distribution of the autoregressive coefficient allows us to obtain closed-form expressions for these terms.

The above suggests an infinite series representation for the aggregated process of the form

$$x_t = \sum_{j=0}^{\infty}\phi_j\nu_{t-j},$$

where $\nu_j \sim N(0,\sigma_\varepsilon^2)$ and $\phi_j = \left(B(p+j,q)/B(p,q)\right)^{1/2}$, $\forall j \in \mathbb{N}$. Note that ν_j inherits the white noise properties of $\varepsilon_{i,t-j}$. Moreover, given Stirling's approximation, the coefficients show a hyperbolic rate of decay with parameter $d = 1-q/2$, that is, $\phi_j \approx j^{-q/2} = j^{d-1}$ as $j \to \infty$.

Once we have proved that the cross-sectional aggregated series can be expressed as a generalized fractional process, Theorem 2 is a direct consequence of Theorem 4.6 in Beran et al. (2013).

Proof of Theorem 3 and Corollary 1

Let $y_t = (1-L)^d x_t$ where x_t is defined as before, then

$$\begin{aligned} E[y_t^2] &= E\left[\left((1-L)^d x_t\right)^2\right] = E\left[\left((1-L)^d \frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_{i,t}\right)^2\right] \\ &= E\left[\frac{1}{N}\left(\sum_{i=1}^{N}(1-L)^d x_{i,t}\right)^2\right] = \frac{1}{N}E\left[\sum_{i=1}^{N}\left((1-L)^d x_{i,t}\right)^2\right], \end{aligned}$$

where the last equality is due to independence across units. Note that the term $(1-L)^d x_{i,t}$ is an ARFIMA(1,$-d$,0); thus the variance of y_t depends on the expected value of the $AR(1)$ coefficient of an $ARFIMA(1,-d,0)$ process.

Let $\gamma_i(k) = E\left[(1-L)^d x_{i,t}(1-L)^d x_{i,t-k}\right]$ be the autocovariance function of $(1-L)^d x_{i,t}$. From Sowell (1992) we know that for $k \in \mathbb{N}$

$$\gamma_i(k|\alpha_i) = \gamma^*(k)\frac{1}{1-\alpha_i^2}\left(F[\{-d+k,1\},1+d+k;\alpha_i] + F[\{-d-k,1\},1+d-k;\alpha_i]-1\right),$$

where

$$\gamma^*(k) = \sigma_\varepsilon^2 \frac{\Gamma(1+2d)}{\Gamma(-d)\Gamma(1+d)}\frac{\Gamma(-d-k)}{\Gamma(1+d-k)},$$

is the autocovariance function of an $I(-d)$ process with innovations with variance σ_ε^2 and $F[\cdot]$ is the hypergeometric function.

Thus,

$$\begin{aligned} \gamma_y(k) &= E\left[\gamma_i(k|\alpha_i)\right] \\ &= E\left[\frac{\gamma^*(k)}{1-\alpha_i^2}\left(F[\{-d+k,1\},1+d+k;\alpha_i]+F[\{-d-k,1\},1+d-k;\alpha_i]-1\right)\right] \\ &= \frac{\gamma^*(k)}{B(p,q)}\left[\int_0^1 (1-x)^{q-2}x^{p-1}F[\{-d+k,1\},1+d+k;x^{\frac{1}{2}}]dx + \right. \\ &\quad \left. \int_0^1 (1-x)^{q-2}x^{p-1}F[\{-d-k,1\},1+d-k;x^{\frac{1}{2}}]dx - \int_0^1 (1-x)^{q-2}x^{p-1}dx\right] \\ &= \frac{\gamma^*(k)}{B(p,q)}\left[B(p,q-1)\left(F_1(k)-1\right) + B(p+\frac{1}{2},q-1)F_2(k)\right], \end{aligned}$$

where

$$
\begin{aligned}
F_1(k) &:= F\left[\left\{1,p,\frac{1-d+k}{2},\frac{-d+k}{2}\right\},\left\{p+q-1,\frac{2+d+k}{2},\frac{1+d+k}{2}\right\},1\right]+\\
&\quad F\left[\left\{1,p,\frac{1-d-k}{2},\frac{-d-k}{2}\right\},\left\{p+q-1,\frac{2+d-k}{2},\frac{1+d-k}{2}\right\},1\right]\\
F_2(k) &:= \frac{-d+k}{1+d+k}*\\
&\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d+k}{2},\frac{2-d+k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d+k}{2},\frac{3+d+k}{2}\right\},1\right]\\
&\quad +\frac{-d-k}{1+d-k}*\\
&\quad F\left[\left\{1,p+\frac{1}{2},\frac{1-d-k}{2},\frac{2-d-k}{2}\right\},\left\{p+q-\frac{1}{2},\frac{2+d-k}{2},\frac{3+d-k}{2}\right\},1\right].
\end{aligned}
$$

Note that in the calculations above we have used

$$
\begin{aligned}
\int_0^1 F[\{a,1\},b;x^{\frac{1}{2}}]x^{p-1}(1-x)^{q-2}dx &= \int_0^1\left[\sum_{i=0}^{\infty}\frac{(a)_i}{(b)_i}x^{\frac{i}{2}}\right]x^{p-1}(1-x)^{q-2}dx\\
=\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\int_0^1 x^{p-1+\frac{i}{2}}(1-x)^{q-2}dx\right] &= \sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}B\left(p+\frac{i}{2},q-1\right)\right].
\end{aligned}
$$

Now,

$$
\begin{aligned}
\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}B\left(p+\frac{i}{2},q-1\right)\right] &= \sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\frac{\Gamma(p+\frac{i}{2})\Gamma(q-1)}{\Gamma(p+q-1+\frac{i}{2})}\right]\\
&= \Gamma(q-1)\sum_{i=0}^{\infty}\left[\frac{(a)_i}{(b)_i}\frac{\Gamma(p+\frac{i}{2})}{\Gamma(p+q-1+\frac{i}{2})}\right]\\
&= \Gamma(q-1)\left(\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}}{(b)_{2i}}\frac{\Gamma(p+i)}{\Gamma(p+q-1+i)}\right]+\right.\\
&\quad \left.\sum_{i=0}^{\infty}\left[\frac{(a)_{2i+1}}{(b)_{2i+1}}\frac{\Gamma(p+\frac{1}{2}+i)}{\Gamma(p+q-\frac{1}{2}+i)}\right]\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \Gamma(q-1)\left(\frac{\Gamma(p)}{\Gamma(p+q-1)}\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}}{(b)_{2i}}\frac{(p)_i}{(p+q-1)_i}\right]+\right.\\
&\quad \left.\frac{\Gamma(p+\frac{1}{2})}{\Gamma(p+q-\frac{1}{2})}\sum_{i=0}^{\infty}\left[\frac{(a)_{2i+1}}{(b)_{2i+1}}\frac{(p+\frac{1}{2})_i}{(p+q-\frac{1}{2})_i}\right]\right)\\
&= B\left(p,q-1\right)\sum_{i=0}^{\infty}\left[\frac{(a)_{2i}(p)_i}{(b)_{2i}(p+q-1)_i}\right]+\\
&\quad B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}\sum_{i=0}^{\infty}\left[\frac{(a+1)_{2i}(p+\frac{1}{2})_i}{(b+1)_{2i}(p+q-\frac{1}{2})_i}\right]\\
&= B\left(p,q-1\right)\sum_{i=0}^{\infty}\left[\frac{(\frac{a}{2})_i(\frac{a+1}{2})_i(p)_i}{(\frac{b}{2})_i(\frac{b+1}{2})_i(p+q-1)_i}\right]+\\
&\quad B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}\sum_{i=0}^{\infty}\left[\frac{(\frac{a+1}{2})_i(\frac{a+2}{2})_i(p+\frac{1}{2})_i}{(\frac{b+1}{2})_i(\frac{b+2}{2})_i(p+q-\frac{1}{2})_i}\right]\\
&= B\left(p,q-1\right)f_1+B\left(p+\frac{1}{2},q-1\right)\frac{a}{b}f_2,
\end{aligned}
$$

where

$$
\begin{aligned}
f_1 &= F\left[\left\{1,p,\frac{a}{2},\frac{a+1}{2}\right\},\left\{p+q-1,\frac{b}{2},\frac{b+1}{2}\right\},1\right],\\
f_2 &= F\left[\left\{1,p+\frac{1}{2},\frac{a+1}{2},\frac{a+2}{2}\right\},\left\{p+q-1,\frac{b+1}{2},\frac{b+2}{2}\right\},1\right],
\end{aligned}
$$

$(z)_i := \frac{\Gamma(z+i)}{\Gamma(z)}$ is the Pochhammer symbol, and noting that $(a)_{2i} = (\frac{1}{2})^{-2i}(\frac{a}{2})_i(\frac{a+1}{2})_i$, $i \in \mathbb{N}$.

For the corollary note that $\gamma_y(k)$ can be written as

$$
\begin{aligned}
\gamma_y(k) &= \frac{\gamma^*(k)}{B(p,q)}\left[-B(p,q-1)+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d+k+i)\Gamma(1+d+k)}{\Gamma(-d+k)\Gamma(1+d+k+i)}\right)B(p+i/2,q-1)\right.\\
&\quad \left.+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d-k+i)\Gamma(1+d-k)}{\Gamma(-d-k)\Gamma(1+d-k+i)}\right)B(p+i/2,q-1)\right].
\end{aligned}
$$

Let

$$\begin{aligned}\tau(k) \quad &:= \quad \frac{1}{B(p,q)}\left[-B(p,q-1)+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d+k+i)\Gamma(1+d+k)}{\Gamma(-d+k)\Gamma(1+d+k+i)}\right)B(p+i/2,q-1)\right.\\ &\left.+\sum_{i=0}^{\infty}\left(\frac{\Gamma(-d-k+i)\Gamma(1+d-k)}{\Gamma(-d-k)\Gamma(1+d-k+i)}\right)B(p+i/2,q-1)\right],\end{aligned}$$

and note that, by Stirling's approximation, for large k and $c>0$, $\Gamma(1+d+ck)/\Gamma(-d+ck)\approx(ck)^{1+2d}$, $\Gamma(-d+ck+i)\Gamma(1+d+ck+i)\approx(ck)^{-1-2d}$ and analogous approximations for the terms in the second series show that

$$\tau(ck) \quad \approx \quad \frac{1}{B(p,q)}\left[-B(p,q-1)+2\sum_{i=0}^{\infty}B(p+i/2,q-1)\right].$$

This, in turn, shows that $\lim_{k\to\infty}\tau(ck)/\tau(k)=1$.

Hence, for large k,

$$\gamma_y(k)=\tau(k)\gamma^*(k)\approx\tau(k)k^{-1-2d},$$

where $\lim_{k\to\infty}\tau(ck)/\tau(k)=1$.

Finally, note that $\sum_{i=0}^{\infty}|\rho_y(k)|=\sum_{i=0}^{\infty}|\gamma_y(k)/\gamma_y(0)|\approx\sum_{i=0}^{\infty}k^{-1-2d}=\zeta(-1-2d)$ where $\zeta(z)$ is the Euler-Riemann zeta function which converges for $z<1$.

CHAPTER 2

FORECASTING LONG MEMORY PROCESSES WITH THE *ARFIMA* MODEL

J. Eduardo Vera-Valdés
Aarhus University and CREATES

Abstract

Most forecasting comparison studies for long memory processes assume that the series are generated by $ARFIMA$ processes. We assess the performance of the $ARFIMA$ model when forecasting long memory series where the long memory generating mechanism may be different from an $ARFIMA$ process. We consider Granger's cross-sectional aggregation, and Parke's error duration model as possible long memory generating mechanisms. We find that $ARFIMA$ models produce similar forecast performance compared to high-order *AR* models at shorter horizons. As the forecast horizon increases, the $ARFIMA$ models tend to dominate in terms of forecast performance. Hence, $ARFIMA$ models are well suited for long horizon forecasts of long memory processes regardless of how the long memory is generated. Additionally, we analyze the forecasting performance of the heterogenous autoregressive model (HAR) which imposes restrictions on high-order *AR* models. We find that the structure enforced by the HAR model produces better long horizon forecasts than *AR* models of the same order, but at the price of inferior short horizon forecasts in some cases.

2.1 Introduction

In the long memory time series literature the $ARFIMA$ class of models remains to be the most popular given its appeal of bridging the gap between the stationary $ARMA$ models and the non-stationary $ARIMA$ model with a unit root. Some effort has been directed to assess the performance of the $ARFIMA$ type of models when forecasting long memory processes.

For instance, Ray (1993) calculates the percentage increase in mean-squared error (MSE) from forecasting $FI(d)$ series with AR models. He argues that the MSE may not increase significantly, particularly when we do not know the true long memory parameter. Crato and Ray (1996) compare the forecasting performance of $ARFIMA$ models against $ARMA$ alternatives and find that $ARFIMA$ models are in general outperformed by $ARMA$ alternatives.

Looking at real data, Martens, van Dijk, and de Pooter (2009) show that for daily realized volatility for forecast horizons of up to twenty days, it seems to be beneficial to use a flexible high-order AR model instead of a parsimonious but stringent fractionally integrated model specification. On the other hand, Barkoulas and Baum (1997) find improvements in forecasting accuracy when fitting $ARFIMA$ models to Eurocurrency returns series, particularly for longer horizons. By allowing for larger data sets of both financial and macro variables, and considering larger forecast horizons, Bhardwaj and Swanson (2006) find that $ARFIMA$ processes generally outperform $ARMA$ alternatives in terms of forecasting performance. Thus, there does not seem to be any consensus regarding the empirical evidence using a forecast metric.

One thing that most forecasting comparison studies have in common is the underlying assumption that long memory is generated by an $ARFIMA$ process. There are two dominant theoretical explanations for the presence of long memory in the time series literature: cross-sectional aggregation of dynamic persistent micro units (Granger, 1980), and that shocks may be of random duration (Parke, 1999). Neither of these sources of long memory imply an $ARFIMA$ specification. Nonetheless, the question of whether an $ARFIMA$ specification serves as a good approximation for forecasting purposes remains.

In this paper, we assess via Monte Carlo simulations the forecast performance of $ARFIMA$ model specifications when the long memory series are generated by other sources than $ARFIMA$ processes. We find that $ARFIMA$ models produce comparable forecast performance as high-order AR models at short and medium forecast horizons. As the forecast horizon increases, the $ARFIMA$ models tend to produce better forecasting performance. Hence, $ARFIMA$ models are well suited for long horizon forecast of long memory regardless of the underlying generating mechanism.

This paper proceeds as follows. In Section 2.2, we present the long memory generating processes to be considered in the simulation study. Section 2.3 describes the design of the Monte Carlo analysis used for the forecasting analysis. Sections 2.4 and 2.5 present and discuss the results from the forecasting analysis, while Section 2.6 concludes.

2.2 Long Memory Generating Processes

In this section, we present the selected processes that generate long memory. All processes considered are long memory in the covariance sense. In contrast to other definitions of long memory, the definition in the covariance sense relates to the rate of decay of the autocorrelations, which the models try to mimic. Thus, the covariance sense is a sensible definition of long memory for forecasting purposes.

2.2.1 The *ARFIMA* Model

As a benchmark, we include in the study the pure $ARFIMA$ process due to Granger and Joyeux (1980), and Hosking (1981). They extended the $ARMA$ model to include fractional dynamics by considering the process

$$\phi(L)(1-L)^d x_t = \theta(L)\epsilon_t, \tag{2.1}$$

where ϵ_t is a white noise process, $d \in (-1/2, 1/2)$, $\phi(L)$ and $\theta(L)$ are polynomials in the lag operator with no common roots, all outside the unit circle. They used the standard binomial expansion to decompose the fractional difference operator $(1-L)^d$ in a series with coefficients $\pi_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$ for $j \in \mathbb{N}$. Using Stirling's approximation, it can be shown that these coefficients decay at a hyperbolic rate, which in turn translates to slowly decaying autocorrelations. Thus, x_t has long memory in the covariance sense.[1]

It is well known that $ARFIMA$ processes are long memory by all definitions typically considered in the literature, and are relatively easy to estimate by Maximum Likelihood. Thus, this has become the canonical construction for modelling and forecasting long memory in the time series literature.

For the Monte Carlo simulations, we consider $ARFIMA(1,d,0)$ processes as a way to incorporate both long and short term dynamics.

[1] In this work, we focus on the definition of long memory in the covariance sense. See among others, Guégan (2005) and Haldrup and Vera-Valdés (2015) for other, often equivalent, definitions of long memory.

2.2.2 Cross-Sectional Aggregation

Granger (1980), in line with the work of Robinson (1978) on autoregressive processes with random coefficients, showed that aggregating $AR(1)$ processes with coefficients sampled from a Beta distribution can produce long memory. He considered N series generated as

$$x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} \quad i = 1, 2, \cdots, N;$$

where $\varepsilon_{i,t}$ is a white noise process with $E[\varepsilon_{i,t}^2] = \sigma_\varepsilon^2 \ \forall i \in \{1, 2, \cdots, N\}, \forall t \in \mathbb{Z}$. Moreover, $\alpha_i^2 \sim \mathscr{B}(\alpha; p, q)$ with $p, q > 1$, and where $\mathscr{B}(\alpha; p, q)$ is the Beta distribution with density given by

$$\mathscr{B}(\alpha; p, q) = \frac{1}{B(p,q)} \alpha^{p-1}(1-\alpha)^{q-1} \quad \text{for} \quad \alpha \in (0,1),$$

with $B(\cdot,\cdot)$ the Beta function. Furthermore, define the cross-sectional aggregated series as

$$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}.$$

Granger showed that, as $N \to \infty$, the autocorrelations of x_t decay at a hyperbolic rate with parameter $d = 1 - q/2$. Thus, x_t has long memory in the covariance sense.

The cross-sectional aggregation result has been extended in different ways, including to allow for general $ARMA$ processes, and to other distributions. See for instance, Oppenheim and Viano (2004), Linden (1999), and Zaffaroni (2004).

Haldrup and Vera-Valdés (2015) show that, although the long memory can be removed by fractional differencing, the resulting series does not belong to the class of linear $ARMA$ processes. The question addressed in this paper is whether an $ARFIMA$ specification is useful for forecasting purposes.

For the Monte Carlo analysis, we generate long memory by cross-sectional aggregation of both $AR(1)$, and $ARMA(1,1)$ processes, the latter as a way to allow more short term dynamics in the specification.

2.2.3 Error Duration Model

Parke (1999) introduced the error duration model where he showed that if the series is the result of the sum of shocks of stochastic duration, then the resulting series would exhibit long memory in the form of hyperbolic decaying autocorrelations.

In particular, let ε_s be a series of $i.i.d.$ shocks with mean zero and finite variance σ^2. Assume that the shock ε_i has a stochastic duration $n_i \geq 0$ time periods, and thus surviving from period i until period $i + n_i$. Let p_k be the probability that event ε_i

survives until period $i+k$, take $g_{i,t}$ to be the indicator function for the event that the error ε_i survives until period t, and define the x_t series as

$$x_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s.$$

Then, if $p_k \sim k^{-2+2d}$ as $k \to \infty$, x_t will have long memory in the covariance sense.

By properly choosing the error survival probabilities, Parke showed that the autocorrelation function will decay at a rate similar to $FI(d)$ processes. However, the resulting series has dichotomic coefficients that do not correspond to a fractional integrated specification, and thus, it is not an $ARFIMA$ process.

We follow Parke's specification in the Monte Carlo simulations and consider error survival probabilities that mimic those of the $FI(d)$ model.

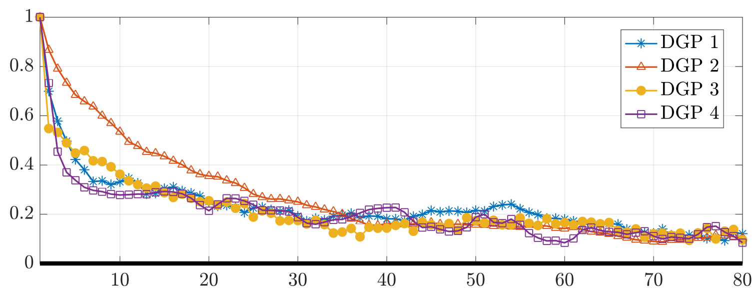
In Table 2.1, we summarize the different long memory generating mechanisms to be analyzed.

Table 2.1. Long Memory Generating Processes

$ARFIMA(p,d,q)$ (DGP 1)	$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t$ $\phi(z) = 1-\phi_1 z - \cdots - \phi_p z^p$ $\theta(z) = 1+\theta_1 z + \cdots + \theta_q z^q$ $(1-L)^d = \sum_{s=0}^{\infty} \frac{\Gamma(s-d)}{\Gamma(-d)\Gamma(s+1)} L^s$
Cross-Sectional Aggregation of $AR(1)$ (DGP 2)	$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}$ $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t}$ $\alpha_i \sim \mathscr{B}(\alpha; p,q);\ p,q>1$
Cross-Sectional Aggregation of $ARMA(1,1)$ (DGP 3)	$x_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{i,t}$ $x_{i,t} = \alpha_i x_{i,t-1} + \varepsilon_{i,t} + \theta\varepsilon_{t-1}$ $\alpha_i \sim \mathscr{B}(\alpha; p,q);\ p,q>1$
Error Duration Model (DGP 4)	$x_t = \sum_{s=-\infty}^{t} g_{s,t}\varepsilon_s$ $g_{s,s+k} = \begin{cases} 0 & \text{w.p. } 1-p_k \\ 1 & \text{w.p. } p_k \end{cases}$ $p_k = k^{2d-2}$

As previously noted, all processes considered in the Monte Carlo analysis are long memory in the covariance sense; that is, they have autocorrelation functions showing hyperbolic decay. This can be seen in Figure 2.1 where we plot the autocorrelation function for the four processes when generating long memory with long memory parameter $d = 0.4$. The figure shows that the autocorrelation function for all processes remain significant at large lags.

Figure 2.1. Autocorrelation function for the four processes considered with $d = 0.4$. The specific parameter specifications chosen for the graphs are presented in Appendix A.



2.3 Monte Carlo Analysis

In this section, we describe the Monte Carlo simulations comparing the forecasting performance of $ARFIMA$ models against $ARMA$ and high-order AR models on long memory series generated by the schemes described in Section 2.2.

2.3.1 Forecast Evaluation

We use the Model Confidence Set (MCS) approach of Hansen et al. (2011) to assess the forecasting performance of the selected models. From an initial set of models, the methodology allows us to obtain the superior set at a given confidence level. In this sense, the MCS is well suited to compare the forecast performance of a large set of competing models.

The MCS algorithm proceeds as follows. From a starting set of competing model, $\mathcal{M}_0$, we search for the set of superior models at forecast horizon h, $\mathcal{M}^*$, defined by

$$\mathcal{M}^* = \{i \in \mathcal{M}_0 \mid E(d^h_{i,j}) \leq 0 \quad \forall j \in \mathcal{M}_0\},$$

where $d^h_{i,j}$ is the loss differential between models i and j.

We obtain $\mathcal{M}^*$ by sequential elimination. For all long memory generating processes, we fit the competing models in the starting set for a sample size T. The models are indexed by $i \in \{1, 2, \dots, m\}$, and the out of sample forecast from model i is denoted by $\hat{y}^i_{T+k}$, $\forall k \in \{1, \dots, h\}$. We rank the models according to their expected loss using one of two loss functions: the mean square error (MSE), $L_{SQ}\left(y_{T+k}, \hat{y}^i_{T+k}\right) = \left(y_{T+k} - \hat{y}^i_{T+k}\right)^2$, and the mean absolute deviation (MAD), $L_{AD}\left(y_{T+k}, \hat{y}^i_{T+k}\right) = \left|y_{T+k} - \hat{y}^i_{T+k}\right|$.

We then define the loss differential between models i and j by

$$d^k_{i,j} = L_M\left(y_{T+k}, \hat{y}^i_{T+k}\right) - L_M\left(y_{T+k}, \hat{y}^j_{T+k}\right),$$

for $= SQ, AD$; $i, j \in \{1, 2, \dots, m\}$.

At each step, we eliminate the worst performing model. We continue with the process until we can not reject the null hypothesis of equal loss differentials for all models in the set; that is,

$$H_0 : E(d^k_{i,j}) \leq 0 \quad \forall i, j \in \mathcal{M}.$$

The null is tested by using either the range statistic, T_R, or the semiquadratic statistic, T_{SQ}, defined by

$$T_R = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{i,j}|}{\left(\widehat{var}(\bar{d}_{i,j})\right)^{1/2}} \qquad T_{SQ} = \sum_{i \neq j} \frac{(\bar{d}_{i,j})^2}{\left(\widehat{var}(\bar{d}_{i,j})\right)^{1/2}}.$$

In the Monte Carlo analysis, we present the proportion of times each model is contained in $\mathcal{M}^*$ for each forecast horizon.

Additionally, as another measure of forecast performance, we compute both the out of sample root mean square error ($RMSE$), and the out of sample root mean absolute deviation ($RMAD$) given by

$$RMSE^i_h = \left(\frac{1}{h}\sum_{k=1}^{h}\left(y_{T+k} - \hat{y}^i_{T+k}\right)^2\right)^{1/2} \quad RMAD^i_h = \left(\frac{1}{h}\sum_{k=1}^{h}\left|y_{T+k} - \hat{y}^i_{T+k}\right|\right)^{1/2},$$

where h and $\hat{y}^i_s$ are defined as above. We report the mean of both $RMSE$ and $RMAD$ across all replications.

2.3.2 Model Selection

As a first step, we use the Bayesian Information Criterion (BIC) to select the number of lags in both the $ARFIMA$ and $ARMA$ models. The validity of the BIC for the class of processes with fractional differencing was proven by Beran, Bhansali, and Ocker (1998). The authors show that for this class of processes, the penalty term must tend to infinity simultaneously with the sample size. Thus, the Akaike Information Criterion is not consistent while the BIC is.

We allow for a maximum of two lags at both components of the $ARFIMA$ model, while the maximum was set to four for the $ARMA$ model. Furthermore, we also allow for a first difference in the $ARMA$ specifications. We use Maximum Likelihood for the estimation of both classes of models with parameter specifications as reported in Appendix A.

Results from the lag selection exercise, presented in Appendix B, show that not many lags are selected for the $ARFIMA$ specification for either component. This suggests that the short term component is not that persistent once we control for the long memory behavior. For the $ARMA$ specification, perhaps not surprisingly, more lags are selected due to the fact that we are not controlling for the long memory behavior via estimation of the fractional memory d. Nonetheless, the maximum number of lags selected by the BIC is two.

Following the results from the lag selection exercise, we present the competing models for the forecasting analysis in Table 2.2. These constitute the starting set, $\mathscr{M}_0$, for the MCS approach explained above.

Table 2.2. Starting Set $\mathscr{M}_0$

$FI(d)$	$ARMA(1,1)$	$HAR(3)$
$ARFIMA(1,d,0)$	$ARMA(2,1)$	$AR(22)$
$ARFIMA(0,d,1)$	$ARMA(1,2)$	$AR(30)$
$ARFIMA(1,d,1)$	$ARMA(3,3)$	$AR(50)$
$ARFIMA(2,d,1)$	$ARMA(4,4)$	$I(1)$

In addition to the preferred models from the lag selection exercise, we also consider high-order AR processes, $AR(30)$ and $AR(50)$. Moreover, given the success of the $HAR(3)$ model of Corsi (2009) on mimicking long memory behavior,[2] we include both the unconstrained $AR(22)$, and the $HAR(3)$ models.

The $HAR(3)$ model is a constrained $AR(22)$ given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + \epsilon_t,$$

where $x_{t-1}^{(f)} = x_{t-1}$, $x_{t-1}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i}$ and, $x_{t-1}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} x_{t-i}$.

[2] See for instance, Andersen, Bollerslev, and Diebold (2007) and Chiriac and Voev (2011).

The HAR specification has been used to model financial data, and reflects that different agents respond to uncertainty at distinct horizons. In this context, the three components of the model seek to capture the daily ($x_t^{(f)}$), weekly ($x_t^{(w)}$), and monthly ($x_t^{(m)}$) levels of uncertainty.

Note that including the $HAR(3)$ model allows us to extend Corsi's (2009) results in several directions. In particular, we make comparisons against a larger set of models; and we use the MCS approach, which is better suited for comparisons between multiple alternatives. Also, we remove the uncertainty regarding the presence of long memory in the data by comparing the performance of the HAR model in simulated long memory series; whereas Corsi used real data.

2.3.3 Monte Carlo Design

The simulations for the forecasting exercise proceed as follows for R replications:

- Generate series of size $T + h$ using the long memory generating processes considered, Section 2.2, Table 2.1. The model calibrations are reported in Table 2.7 in Appendix A.
- Fit by Maximum Likelihood the competing models in the starting set $\mathcal{M}_0$, Table 2.2, for a sample size T.
- Construct forecasts from each model for horizons $h \in \{5, 10, 30, 50, 100, 300\}$.
- Determine the MCS and compute the $RMSE$ and $RMAD$.

After the R replications, we report the proportion of times each model is contained in the MCS, and the mean values of $RMSE$ and $RMAD$, for each forecast horizon.

Throughout, we use a large sample size of $T = 1,000$ to reduce estimation error. We consider values of the long memory parameter in the $(0, 1/2)$ range to produce stationary series and avoid having to take first differences. Furthermore, given the rise of climate econometrics studies keen on producing really long forecasts, we consider it relevant to evaluate forecast performances to horizons as far as $h = 300$, which correspond to twenty-five years of monthly forecasts.

2.4 Results

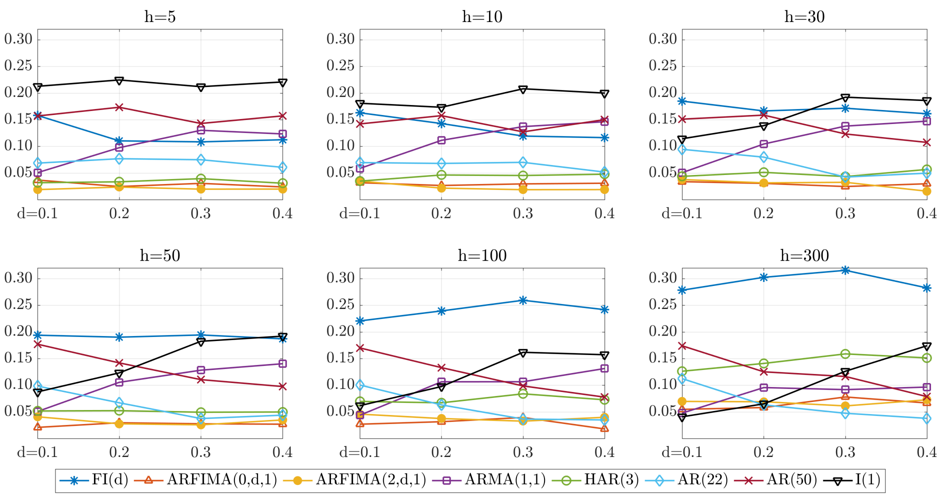
2.4.1 *DGP* 1

As a benchmark, we present in Table 2.3 and Figure 2.2 the results from the Monte Carlo analysis for an $ARFIMA(1, d, 0)$ process, DGP 1, for $d = 0.3$. The table and figure present the results for the MAD loss function and the T_R statistic. Throughout, for ease of exposition, we focus on the MAD loss function; nonetheless, tables using the MSE loss function, reported in Appendix C, show similar results.

Table 2.3. Mean of the $RMAD$ and proportion of times the model is in the MCS using the MAD loss function and the T_R statistic at a 95% confidence level.

DGP 1	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	0.937	0.109	0.964	0.118	0.984	0.171	0.988	**0.193**	0.995	**0.259**	1.004	**0.315**
$ARFIMA(1,d,0)$	**0.933**	0.029	0.962	0.028	0.983	0.026	0.988	0.034	**0.994**	0.035	**1.003**	0.055
$ARFIMA(0,d,1)$	**0.933**	0.031	**0.961**	0.030	**0.982**	0.025	**0.987**	0.028	**0.994**	0.039	**1.003**	0.078
$ARFIMA(1,d,1)$	0.935	0.009	0.963	0.011	0.983	0.018	0.988	0.028	**0.994**	0.025	**1.003**	0.053
$ARFIMA(2,d,1)$	0.935	0.020	0.963	0.019	0.984	0.033	0.989	0.026	0.995	0.033	1.004	0.061
$ARMA(1,1)$	0.944	0.130	0.975	0.137	0.995	0.138	0.999	0.128	1.002	0.107	1.007	0.092
$ARMA(2,1)$	0.937	0.023	0.966	0.027	0.988	0.013	0.993	0.025	0.999	0.024	1.006	0.045
$ARMA(1,2)$	0.938	0.036	0.968	0.032	0.990	0.039	0.995	0.037	1.000	0.034	1.006	0.043
$ARMA(3,3)$	0.937	0.032	0.967	0.034	0.988	0.026	0.993	0.019	0.999	0.025	1.007	0.048
$ARMA(4,4)$	0.938	0.040	0.967	0.042	0.988	0.035	0.993	0.037	0.999	0.030	1.007	0.063
$HAR(3)$	0.936	0.039	0.965	0.045	0.986	0.043	0.992	0.049	0.998	0.084	1.006	0.159
$AR(22)$	0.939	0.075	0.967	0.070	0.988	0.042	0.992	0.038	0.998	0.037	1.005	0.047
$AR(30)$	0.941	0.073	0.970	0.072	0.989	0.076	0.993	0.067	0.998	0.060	1.005	0.071
$AR(50)$	0.948	0.143	0.976	0.127	0.994	0.123	0.997	0.111	1.000	0.099	1.006	0.116
$I(1)$	1.036	**0.212**	1.076	**0.208**	1.113	**0.193**	1.125	0.183	1.151	0.162	1.200	0.126

Figure 2.2. Proportion of times the top performing models are in the MCS at a 95% confidence level when forecasting DGP 1 with different degrees of memory at several horizons.



We can see that $ARFIMA$ models are the preferred specification for all forecast horizons measured by the $RMAD$ criterion, which is not surprising given that DGP 1 is an $ARFIMA$ process. Furthermore, the $I(1)$ model seems to give the worst performance by the $RMAD$ criterion. Noting that the $RMAD$ is a measure across all replications, it suggests that when the $I(1)$ is not in the confidence set, its forecasts behave badly.

Turning to the MCS criterion, note that the no-change $I(1)$ model gives the best forecast performance for short horizons. Nonetheless, we find that high-order AR and $ARMA$ models perform quite well when forecasting a true $ARFIMA$ process for short forecast horizons. In particular, the $AR(50)$ and $ARMA(1,1)$ give better performance than $ARFIMA$ specifications for forecast horizons $h = 5$, and $h = 10$. Yet, for larger forecast horizons, the $FI(d)$ model is the preferred one, and its relative performance increases with the forecast horizon. The superior performance of the $FI(d)$ model compared to the correct $ARFIMA(1,d,0)$ specification may be explained given the small value of the autoregressive coefficient. It suggests that the $FI(d)$ model seems to capture enough information for forecasting purposes in the long horizon once the short memory component fades away.[3] The table is in line with the findings of previous studies on forecasting long memory generated by $ARFIMA$ processes for short and medium forecast horizons, while extending the analysis to larger forecast horizons and by the inclusion of the MCS criterion.

Furthermore, Figure 2.2 allows us to further contrast the performance of high-order AR models and $ARFIMA$ models.[4] The figure shows that for $h = 5$ and $h = 10$, the $AR(50)$ produces better or similar forecast performance than the $FI(d)$ model. As the forecast horizon increases, the $FI(d)$ models tend to lead in forecast performance. Thus, it seems to indicate that, for DGP 1, $ARFIMA$ models are well suited to make forecasts for medium and long horizons, while high-order AR models work well at short forecast horizons.

Finally, the figure allows us to compare the $HAR(3)$ model against the $AR(22)$ model. Note the crossing in preferred model between the $AR(22)$ and $HAR(3)$ models as the forecast horizon and degree of memory increase. The figure shows that for $h = 5$, the $AR(22)$ model is always on top of the $HAR(3)$ model. As the forecast horizon increases, the preferred model changes from the $AR(22)$ to the $HAR(3)$ model for higher degrees of memory. For instance, for $h = 50$, the crossing happens at $d = 0.3$; while for $h = 100$, the cross occurs sooner at $d = 0.2$. That is, the structure imposed by the $HAR(3)$ specification improves forecasting performance for higher degrees of memory and for the larger forecast horizons, at the cost of lower performance at small

[3] Results allowing more short-term dynamics are presented in Appendix D.

[4] For ease of exposition, we present a subset of the top performing models in the figures; nonetheless, we present plots with all competing models in Appendix E.

horizons.

Overall, Table 2.3 and Figure 2.2 extend the findings of previous studies on forecasting long memory when the long memory is generated by $ARFIMA$ processes. That is, high-order AR models are good alternatives for shorter forecast horizons; while $ARFIMA$ models are better suited for large forecast horizons. Moreover, the constraints imposed by the HAR model improve forecasting performance over the unconstrained same-order AR model only for higher degrees of memory and longer forecast horizons.

2.4.2 *DGP* 2

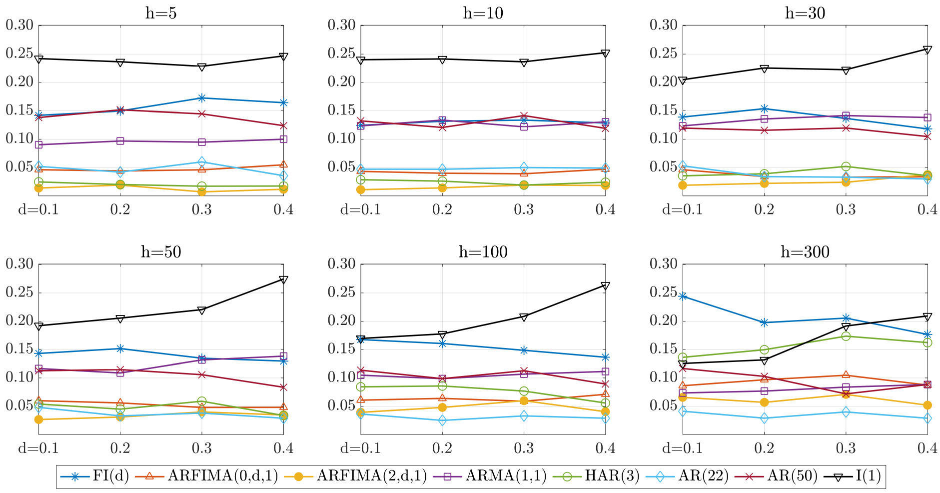
Results from the Monte Carlo analysis for the cross-sectional aggregated $AR(1)$ processes, DGP 2, are presented in Table 2.4. The table presents the results for the MAD loss function and the T_R statistic with long memory parameter $d = 0.3$.

Table 2.4. Mean of the $RMAD$ and proportion of times the model is in the MCS using the MAD loss function and the T_R statistic at a 95% confidence level.

DGP 2	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.027	0.172	1.088	0.134	1.161	0.136	1.192	0.135	1.228	0.149	1.267	**0.206**
$ARFIMA(1,d,0)$	**1.019**	0.036	**1.084**	0.034	**1.159**	0.037	**1.191**	0.037	**1.227**	0.038	**1.266**	0.062
$ARFIMA(0,d,1)$	1.020	0.046	1.085	0.039	**1.159**	0.032	**1.191**	0.047	**1.227**	0.059	**1.266**	0.105
$ARFIMA(1,d,1)$	**1.019**	0.020	**1.084**	0.028	1.161	0.048	1.192	0.045	**1.227**	0.041	**1.266**	0.072
$ARFIMA(2,d,1)$	**1.019**	0.007	1.086	0.019	1.164	0.024	1.196	0.040	1.230	0.060	1.268	0.071
$ARMA(1,1)$	1.029	0.095	1.097	0.121	1.184	0.142	1.216	0.132	1.243	0.107	1.275	0.084
$ARMA(2,1)$	1.022	0.032	1.089	0.034	1.172	0.026	1.205	0.016	1.235	0.028	1.272	0.049
$ARMA(1,2)$	1.026	0.024	1.093	0.026	1.178	0.036	1.212	0.035	1.241	0.033	1.274	0.037
$ARMA(3,3)$	1.026	0.024	1.092	0.028	1.173	0.022	1.205	0.026	1.237	0.030	1.275	0.054
$ARMA(4,4)$	1.024	0.025	1.090	0.028	1.171	0.023	1.204	0.023	1.235	0.029	1.273	0.060
$HAR(3)$	1.021	0.017	1.087	0.019	1.168	0.052	1.203	0.059	1.236	0.077	1.275	0.173
$AR(22)$	1.023	0.060	1.089	0.050	1.168	0.033	1.201	0.038	1.233	0.033	1.271	0.040
$AR(30)$	1.025	0.070	1.092	0.063	1.171	0.047	1.203	0.042	1.234	0.037	1.271	0.035
$AR(50)$	1.032	0.145	1.097	0.142	1.177	0.120	1.208	0.105	1.237	0.112	1.273	0.071
$I(1)$	1.075	**0.227**	1.164	**0.235**	1.275	**0.222**	1.326	**0.220**	1.387	**0.208**	1.479	0.191

Focusing on the MCS criterion, note that the $I(1)$ model is contained in the MCS the most for forecasts horizons up to 100, followed by the $AR(50)$, $ARMA(1,1)$, and $FI(d)$ models. For the largest horizon, $h = 300$, the $FI(d)$ model is the one contained in the MCS the most. Nonetheless, the $I(1)$ model still appears in the MCS quite often even though $d = 0.3$ and hence the series is stationary by construction. Moreover, the $HAR(3)$ model replaces the $ARMA(1,1)$ model in the top three at the largest forecast horizon. This finding is particularly compelling when compared against the

Figure 2.3. Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP* 2 with different degrees of memory at several horizons.



unconstrained $AR(22)$. The additional structure imposed by the HAR model on the autoregressive coefficients works well for really large horizons, while providing slightly worse forecast performance for shorter horizons. Furthermore, while the $AR(50)$ behaves well for the smaller forecast horizon, its performance decays as the horizon increases.

Looking at the $RMAD$ criterion, note that the preferred models for all forecast horizons always belong to the $ARFIMA$ class of models. On the other hand, the $I(1)$ model is the worst performing according to this criterion, which seems to suggest poor average forecast performance.

We can see the effect that the degree of long memory has on the results in Figure 2.3. We plot the proportion of times the models are contained in the MCS for different degrees of memory, for all forecast horizons.

The figure extends the findings in Table 2.4. It shows the good performance of the $ARMA(1,1)$ and $AR(50)$ models for short and medium forecast horizons, providing similar results to the $FI(d)$ specification. Furthermore, when the forecast horizon is large, the figure shows the increase in relative forecast performance of the $FI(d)$ model. Also, the plot shows the good performance of the $HAR(3)$ model for really large forecast horizons. In particular, while the performance in small forecast horizons is inferior in comparison to the unconstrained $AR(22)$, the constrains seem to introduce the additional structure needed for making good long horizon forecasts.

Overall, Table 2.4 and Figure 2.3 indicate that the $ARFIMA$ class of models are a viable alternative for making forecasts when working with long memory series generated by DGP 2, particularly when the forecast horizon is large. On the short memory alternatives, the $AR(50)$ and $ARMA(1,1)$ models are well suited for smaller forecasting periods, while the $HAR(3)$ model starts to show good performance at larger horizons.

2.4.3 *DGP* 3

In Table 2.5, we present the results from the Monte Carlo analysis for the cross-sectional aggregated $ARMA(1,1)$ processes, DGP 3, for long memory parameter $d = 0.3$.

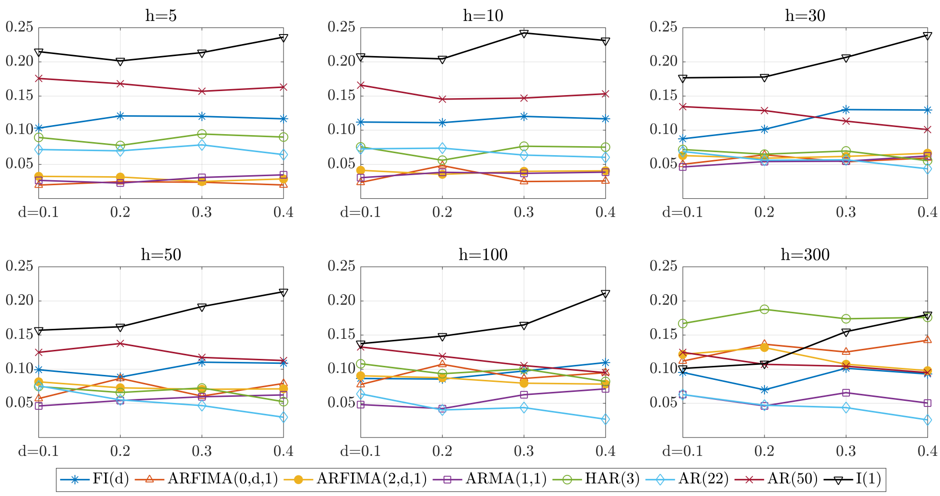
Table 2.5. Mean of the $RMAD$ and proportion of times the model is in the MCS using the MAD loss function and the T_R statistic at a 95% confidence level.

DGP 3	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	0.912	0.120	0.940	0.120	0.968	0.130	**0.979**	0.110	**0.989**	0.097	**1.005**	0.101
$ARFIMA(1,d,0)$	**0.909**	0.019	**0.939**	0.023	**0.967**	0.019	**0.979**	0.020	**0.989**	0.028	**1.005**	0.042
$ARFIMA(0,d,1)$	**0.909**	0.024	**0.939**	0.025	0.968	0.054	**0.979**	0.061	**0.989**	0.086	**1.005**	0.125
$ARFIMA(1,d,1)$	0.915	0.021	0.945	0.017	0.978	0.038	0.992	0.058	1.006	0.069	1.019	0.097
$ARFIMA(2,d,1)$	0.916	0.025	0.946	0.040	0.980	0.062	0.994	0.071	1.009	0.079	1.022	0.107
$ARMA(1,1)$	0.911	0.031	0.940	0.037	0.972	0.055	0.985	0.060	0.995	0.063	1.009	0.066
$ARMA(2,1)$	0.911	0.013	0.940	0.007	0.971	0.019	0.984	0.020	0.995	0.025	1.009	0.044
$ARMA(1,2)$	0.911	0.007	0.940	0.010	0.971	0.012	0.984	0.014	0.995	0.015	1.009	0.026
$ARMA(3,3)$	0.912	0.049	0.941	0.047	0.972	0.039	0.984	0.040	0.994	0.047	1.009	0.054
$ARMA(4,4)$	0.912	0.061	0.941	0.064	0.972	0.064	0.984	0.056	0.995	0.052	1.010	0.055
$HAR(3)$	0.910	0.094	**0.939**	0.076	0.971	0.070	0.984	0.072	0.995	0.100	1.011	**0.174**
$AR(22)$	0.915	0.078	0.944	0.064	0.972	0.057	0.983	0.047	0.993	0.044	1.009	0.044
$AR(30)$	0.917	0.088	0.945	0.082	0.973	0.065	0.984	0.065	0.994	0.066	1.008	0.072
$AR(50)$	0.925	0.157	0.954	0.147	0.981	0.113	0.989	0.117	0.996	0.105	1.009	0.104
$I(1)$	1.027	**0.214**	1.053	**0.242**	1.096	**0.207**	1.117	**0.192**	1.142	**0.165**	1.195	0.155

We see that the models that minimize the $RMAD$ criteria are mainly $ARFIMA$ specifications. In particular, the $RMAD$ values are close across low-order $ARFIMA$ specifications. For the $ARFIMA(1,d,1)$ and $ARFIMA(2,d,1)$ models, the $RMAD$ criterion is slightly worse, and thus suggesting that some overfitting may be occurring. Furthermore, once again, the $I(1)$ model gives the worst performance according to the $RMAD$ criterion.

Turning to the MCS criterion, note first that the no-change model $I(1)$ is contained in the MCS the most for horizons up to 100. Yet, the $HAR(3)$ seems to provide good

Figure 2.4. Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP* 3 with different degrees of memory at several horizons.



forecast performance for larger horizons, being the one contained in the *MCS* the most for $h = 300$. The performance of the $HAR(3)$ model is particularly noteworthy when compared with the performance of unconstrained high-order *AR* alternatives at long horizons. Furthermore, note the good performance of the $AR(50)$ when the forecast horizon is small, it beats all fractional differenced alternatives for $h = 5$ and $h = 10$; nonetheless, its performance decays as the forecast horizon increases.

For the fractional difference specifications, note the increase in forecast performance of the $ARFIMA(0,d,1)$ model as the forecast horizon increases. For $h = 300$, it is the second preferred model. Moreover, it increases in relative performance against the pure $FI(d)$ model as the forecast horizon increases. This points to the gains that can be made by controlling the additional short memory dynamics introduced in *DGP* 3 relative to *DGP* 2 with a higher order *ARFIMA* model. Thus, the increase in complexity seems to be beneficial at longer forecast horizons.

We present the proportion of times each model is contained in the *MCS* for different values of the theoretical long memory parameter in Figure 2.4.

The figure shows that the good performance of the $AR(50)$ model at shorter horizons extend to all degrees of memory. As the forecast horizon increases, fractional differenced alternatives start to improve in relative terms. Thus, the high-order *AR* and *ARFIMA* models are good complementary candidates for forecasting *DGP* 3 at short and long forecast horizons, respectively.

Finally, the figure shows that the good performance of the $HAR(3)$ model at the longest forecast horizon extends to all the degrees of memory considered, while providing similar forecasting performance than the $AR(22)$ at shorter horizons.

2.4.4 *DGP* 4

Table 2.6 presents the results from the Monte Carlo analysis for processes generated using the error duration model, *DGP* 4, for long memory parameter $d = 0.3$.

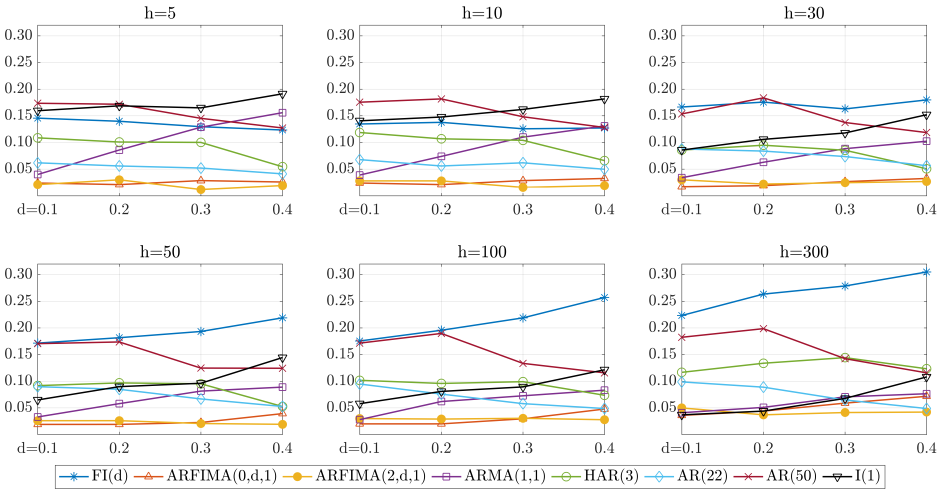
Table 2.6. Mean of the $RMAD$ and proportion of times the model is in the MCS using the MAD loss function and the T_R statistic at a 95% confidence level.

DGP 4	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.097	0.130	1.123	0.126	1.131	**0.162**	1.133	**0.193**	1.136	**0.217**	1.140	**0.277**
$ARFIMA(1,d,0)$	1.075	0.038	1.105	0.037	1.120	0.041	1.124	0.035	1.130	0.042	1.136	0.063
$ARFIMA(0,d,1)$	1.074	0.028	1.109	0.028	1.130	0.027	1.137	0.023	1.145	0.029	1.156	0.059
$ARFIMA(1,d,1)$	**1.067**	0.012	**1.100**	0.012	**1.118**	0.009	**1.123**	0.021	**1.129**	0.021	**1.135**	0.037
$ARFIMA(2,d,1)$	**1.067**	0.012	**1.100**	0.016	**1.118**	0.025	**1.123**	0.021	**1.129**	0.031	**1.135**	0.041
$ARMA(1,1)$	1.077	0.129	1.109	0.110	1.125	0.089	1.129	0.082	1.133	0.073	1.137	0.071
$ARMA(2,1)$	1.074	0.031	1.108	0.040	1.125	0.043	1.128	0.042	1.133	0.043	1.137	0.050
$ARMA(1,2)$	1.070	0.018	1.103	0.020	1.120	0.024	1.125	0.030	1.130	0.030	1.136	0.044
$ARMA(3,3)$	1.069	0.023	1.102	0.025	1.119	0.033	1.124	0.033	1.130	0.040	1.136	0.065
$ARMA(4,4)$	1.070	0.043	1.103	0.033	1.119	0.039	1.124	0.050	1.130	0.047	**1.135**	0.079
$HAR(3)$	1.076	0.100	1.106	0.104	1.121	0.086	1.126	0.095	1.131	0.099	1.137	0.145
$AR(22)$	1.077	0.052	1.108	0.062	1.122	0.074	1.126	0.066	1.131	0.057	1.136	0.065
$AR(30)$	1.078	0.074	1.109	0.077	1.124	0.094	1.126	0.089	1.131	0.085	1.136	0.096
$AR(50)$	1.085	0.146	1.116	0.149	1.131	0.138	1.133	0.125	1.134	0.134	1.137	0.143
$I(1)$	1.212	**0.164**	1.261	**0.161**	1.298	0.116	1.309	0.095	1.324	0.089	1.371	0.068

We can see from the table that the $ARFIMA$ class of models provide the best performance measured by the $RMAD$ criterion for all forecast horizons, while the best performance by the MCS criterion for $h = 30$ and bigger. That is, the $FI(d)$ model is the one contained the most in the MCS for medium and long forecast horizons, while remaining competitive at shorter horizons.

Focusing on short memory alternatives, note the relatively good performance of the $AR(50)$ model for all forecast horizons. The $AR(50)$ model is always in the top three. Thus, even though the $FI(d)$ model is the clear winner as the forecast horizon increases, high-order AR models can produce good forecasts for *DGP* 4. Moreover, contrasting the performance of the $HAR(3)$ model against the $AR(22)$ model, the table shows the gains in performance of imposing some structure into the higher-order AR

Figure 2.5. Proportion of times the top performing models are in the *MCS* at a 95% confidence level when forecasting *DGP* 4 with different degrees of memory at several horizons.



models when the forecast horizon is large.

Figure 2.5 presents the proportion of times the models are contained in the *MCS* when forecasting *DGP* 4 for different degrees of memory. The figure shows the relative performance increase of the $FI(d)$ model over high-order *AR* models as both the degree of memory and the forecast horizon increase. Also, note that the $HAR(3)$ model is always on top, if slightly for medium horizons, of the unconstrained $AR(22)$ model.

Overall, Table 2.6 and Figure 2.5 show compelling evidence in favor of using the *ARFIMA* model to make forecasts of processes generated by the error duration model, *DGP* 4.

2.5 Discussion

The results from the Monte Carlo simulation can be further analyzed in the context of the bias-variance trade-off typically studied in regression analysis.

All processes considered in this paper are long memory in the covariance sense, see Figure 2.1. Hence, the models are fitted to capture the information contained in the autocorrelation function and use it for forecasting purposes. In other words, the models select $\{a_i\}_{i=0}^{T}$ in the representation $x_t = a_0 + \sum_{i=1}^{k} a_i x_{t-i}$, where k is the order of the autoregressive representation, with the aim of replicating the autocorrelation function.

$ARFIMA$ and $ARMA$ models differ in terms of the rate of decay of the coefficients a_i. $ARFIMA$ models impose a hyperbolic rate by the fractional differencing operator $(1-L)^d$, see equation 2.1. Thus, $ARFIMA$ models need just one parameter to establish the infinite list of coefficients, and are hence of low variance. Nonetheless, the uncertainty surrounding the estimation of the long memory parameter may introduce some bias.

As an alternative, high-order AR models are more flexible by choosing each coefficient separately. Hence, they can reduce the bias of the coefficients assigned, but suffer from increased variance given the number of parameters estimated. This is particularly important in the scenario of having small samples for estimation, something we abstract from in this study. Given the uncertainty associated to estimating more parameters, we would expect the performance of high-order AR models to deteriorate in short series. Yet, as the Monte Carlo analysis shows, this flexibility can produce good forecasts at shorter horizons, particularly when the degree of memory is small. Nonetheless, AR models lose forecasting power as the forecast horizon gets larger. We could increase the order of the autoregressive process to produce better long horizon forecasts, but the estimation becomes unstable.

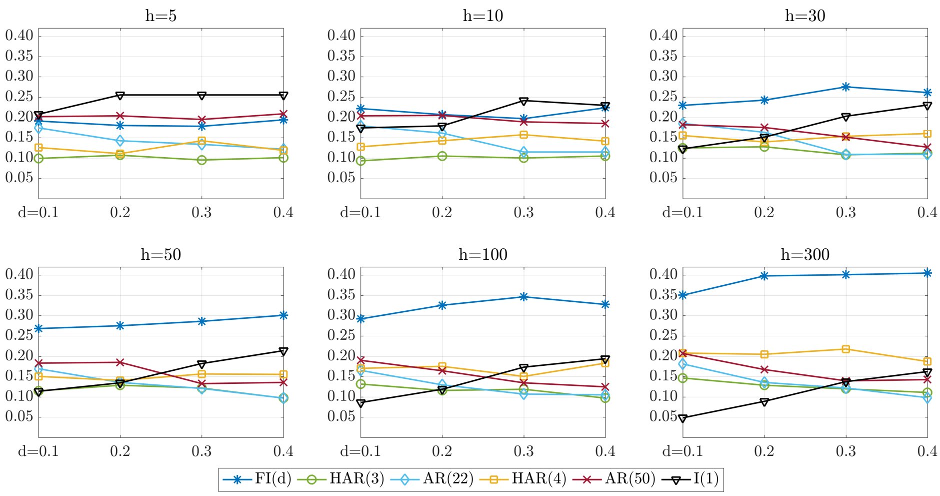
In this context, HAR models are a compromise between the rigid $ARFIMA$ and flexible high-order AR model specifications. They incorporate high-order autoregressive specifications while greatly restricting the number of parameters to be estimated. This arrangement allows the HAR model to provide similar forecast performance at medium forecast horizons as same-order unrestricted AR models, while providing better long horizon forecasts. Yet, HAR models suffer a forecast performance loss at shorter horizons. This can be seen better in Figure 2.6 where we show the average number of times two high-order AR processes and their comparable HAR specifications are contained in the MCS when forecasting DGP 1. In particular, we show a $HAR(4)$ given by

$$x_t = a_0 + a_1 x_{t-1}^{(f)} + a_2 x_{t-1}^{(w)} + a_3 x_{t-1}^{(m)} + a_4 x_{t-1}^{(b)} + \epsilon_t,$$

where $x_{t-1}^{(f)} = x_{t-1}$, $x_{t-1}^{(w)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i}$, $x_{t-1}^{(m)} = \frac{1}{22}\sum_{i=1}^{22} x_{t-i}$, and $x_{t-1}^{(b)} = \frac{1}{50}\sum_{i=1}^{50} x_{t-i}$. Note that the $HAR(4)$ model is a constrained $AR(50)$.

The figure displays the similar performance between constrained and unconstrained autoregressive processes of the same order for medium forecast horizons. Furthermore, it shows the increase in relative performance for the constrained versions at larger forecast horizons. Nonetheless, this increase in forecast performance comes at the price of inferior performance at shorter horizons. In particular, for $h = 5$ and

Figure 2.6. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 1 with different degrees of memory at several horizons. For the plots, the starting set contains only the six models shown.



$h = 10$, the unconstrained *AR* models give better performance than equivalent order *HAR* alternatives for all degrees of memory.[5]

The bias-variance trade-off has been a topic of great interest in the literature of regressions with a high number of covariates. It thus would be compelling to adapt shrinkage and sparse methods to lag selection in the context of long memory forecasting. This line of inquiry is left open for future research.

2.6 Conclusions

This paper evaluates the forecasting performance of *ARFIMA* models when the memory is generated from sources different from the *ARFIMA* model.

We find that high-order *AR* models produce comparable forecasts as *ARFIMA* models at shorter horizons. As the forecast horizon increases, the *ARFIMA* models tend to dominate in terms of forecast performance. Hence, *ARFIMA* models are well suited for long horizon forecasts of long memory, regardless of the generating mechanism. In particular, we find that if the long memory is generated by the error duration model, the $FI(d)$ model produces the best forecast performance at medium

[5] Appendix F shows that this result extends to the other *DGP*s considered.

and large horizons for all degrees of memory, while remaining competitive at shorter horizons.

Additionally, by making a compromise between flexibility and complexity, we find that the structure imposed by the HAR model induces a tradeoff in forecast performance at different forecast horizons. In other words, the HAR model produces better long horizon forecasts, similar medium horizon forecasts, and similar or inferior short horizon forecasts, than same-order AR model specifications.

Acknowledgments

I would like to thank Niels Haldrup for all the insightful comments, the paper greatly improved because of him. The paper was written in part while on a visiting stay at Erasmus University Rotterdam. I would like to thank the Faculty and Administrative Staff at the Erasmus School of Economics for all their help; in particular to Michel van der Wel and Dick van Dijk for the interesting conversations. The author acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

2.7 References

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. Review of Economics and Statistics 89 (4), 701–720.

Barkoulas, J. T., Baum, C. F., 1997. Fractional Differencing Modeling and Forecasting of Eurocurrency Deposit Rates. The Journal of Financial Research 20 (3), 355–372.

Beran, J., Bhansali, R. J., Ocker, D., 1998. On unified model selection for stationary and non stationary short and long memory autoregressive processes. Biometrika 85 (4), 921–934.

Bhardwaj, G., Swanson, N. R., 2006. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Journal of Econometrics 131 (1-2), 539–578.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26 (6), 922–947.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174–196.

Crato, N., Ray, B. K., 1996. Model selection and forecasting for long-range dependent processes. Journal of Forecasting 15 (2), 107–125.

Granger, C. W., 1980. Long memory relationships and the aggregation of dynamic models. Journal of econometrics 14, 227–238.

Granger, C. W., Joyeux, R., 1980. An Introduction To Long-Memory Time Series Models and Fractional Differencing. Journal of Time Series Analysis 1 (1), 15–29.

Guégan, D., 2005. How can we define the concept of long memory? An econometric survey. Econometric Reviews 24 (2), 113–149.

Haldrup, N., Vera-Valdés, J., 2015. Long Memory, Fractional Integration, and Cross-Sectional Aggregation. CREATES Research Paper 59.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Hosking, J. R. M., 1981. Fractional Differencing. Biometrika 68 (1), 165.

Linden, M., 1999. Time series properties of aggregated AR (1) processes with uniformly distributed coefficients. Economics Letters 64, 31–36.

Martens, M., van Dijk, D., de Pooter, M., 2009. Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. International Journal of Forecasting 25 (2), 282–303.

Oppenheim, G., Viano, M. C., 2004. Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results. Journal of Time Series Analysis 25 (3), 335–350.

Parke, W., 1999. What is fractional integration? Review of Economics and Statistics 81 (4), 632–638.

Ray, B. K., 1993. Modeling Long-Memory Processes for Optimal Long-Range Prediction. Journal of Time Series Analysis 14 (5), 511–525.

Robinson, P. M., 1978. Statistical inference for a random coefficient autoregressive model. Scandinavian Journal of Statistics 5 (3), 163–168.

Zaffaroni, P., 2004. Contemporaneous aggregation of linear dynamic models in large economies. Journal of Econometrics 120 (1), 75–102.

2.8 Appendix

A Parameters

Overall	$\varepsilon_t \sim i.i.d.N(0,1)\ \forall t$ $T = 1000$; $R = 1000$
DGP 1	$\phi_1 = 0.2$
DGP 2	$N = 10,000$; $p = 1.4$
DGP 3	$N = 10,000$; $p = 1.4$; $\theta = 0.5$
DGP 4	$p_k = (\Gamma(k+d)\Gamma(2-d))/(\Gamma(k+2-d)\Gamma(d))$

Table 2.7. Parameters for the Monte Carlo simulations

B Lag Selection Exercise

Model	d	*ARFIMA*		*ARIMA*	
		AIC	*BIC*	*AIC*	*BIC*
DGP 2	0.2	(1,1)	(1,0)	(4,3)	(2,1)
		[0.21]	[0.38]	[0]	[0]
	0.4	(2,1)	(1,0)	(2,1)	(2,1)
		[0.10]	[0.46]	[0]	[0]
DGP 3	0.2	(0,1)	(0,1)	(3,3)	(1,1)
		[0.38]	[0.40]	[0]	[0]
	0.4	(0,1)	(0,1)	(4,4)	(1,1)
		[0.45]	[0.46]	[0]	[0]
DGP 4	0.2	(0,1)	(0,1)	(3,4)	(1,1)
		[0.10]	[0.10]	[0]	[0]
	0.4	(0,1)	(0,1)	(4,4)	(1,2)
		[0.27]	[0.26]	[0]	[0]

Table 2.8. Lag selection for the *AR* and *MA* components. We show the preferred model for each criteria from 1,000 replications using a sample size of 1,000. Below, the mean of the associated order of integration estimated is presented.

C *MSE* Loss Function

Table 2.9. Mean of the $RMSE$ and proportion of times the model is in the MCS using the SQ loss function and the T_R statistic at a 95% confidence level.

DGP 1	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.074	0.096	1.148	0.125	1.204	0.174	1.218	**0.214**	1.236	**0.257**	1.261	**0.325**
$ARFIMA(1,d,0)$	1.069	0.025	1.144	0.033	1.204	0.028	1.217	0.032	**1.234**	0.025	**1.259**	0.051
$ARFIMA(0,d,1)$	**1.067**	0.026	**1.143**	0.024	**1.202**	0.027	**1.216**	0.033	**1.234**	0.030	**1.259**	0.069
$ARFIMA(1,d,1)$	1.071	0.013	1.146	0.012	1.204	0.019	1.218	0.023	1.235	0.030	**1.259**	0.042
$ARFIMA(2,d,1)$	1.072	0.019	1.147	0.024	1.206	0.023	1.221	0.027	1.238	0.038	1.261	0.066
$ARMA(1,1)$	1.089	0.142	1.171	0.155	1.233	0.138	1.242	0.117	1.252	0.111	1.267	0.081
$ARMA(2,1)$	1.075	0.019	1.152	0.024	1.215	0.014	1.229	0.016	1.245	0.021	1.266	0.047
$ARMA(1,2)$	1.078	0.032	1.157	0.034	1.221	0.038	1.234	0.043	1.247	0.038	1.265	0.042
$ARMA(3,3)$	1.076	0.033	1.153	0.025	1.214	0.033	1.229	0.035	1.245	0.029	1.267	0.041
$ARMA(4,4)$	1.079	0.048	1.155	0.039	1.214	0.037	1.228	0.038	1.245	0.038	1.268	0.066
$HAR(3)$	1.074	0.030	1.150	0.039	1.211	0.054	1.227	0.050	1.244	0.091	1.265	0.161
$AR(22)$	1.080	0.075	1.156	0.076	1.214	0.044	1.228	0.035	1.243	0.030	1.264	0.026
$AR(30)$	1.087	0.067	1.164	0.060	1.218	0.057	1.230	0.055	1.244	0.056	1.264	0.054
$AR(50)$	1.100	0.147	1.175	0.120	1.230	0.119	1.240	0.105	1.250	0.088	1.266	0.089
$I(1)$	1.303	**0.230**	1.414	**0.211**	1.521	**0.196**	1.555	0.179	1.629	0.159	1.769	0.124

Table 2.10. Mean of the $RMSE$ and proportion of times the model is in the MCS using the SQ loss function and the T_R statistic at a 95% confidence level.

DGP 2	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.281	0.174	1.455	0.152	1.670	0.138	1.765	0.146	1.877	0.156	2.000	**0.216**
$ARFIMA(1,d,0)$	**1.264**	0.028	**1.444**	0.031	**1.667**	0.036	**1.763**	0.039	**1.875**	0.037	**1.999**	0.048
$ARFIMA(0,d,1)$	1.268	0.035	1.447	0.029	**1.667**	0.041	**1.763**	0.053	**1.875**	0.064	**1.999**	0.091
$ARFIMA(1,d,1)$	**1.264**	0.024	1.446	0.035	1.672	0.052	1.766	0.041	1.876	0.048	**1.999**	0.067
$ARFIMA(2,d,1)$	1.265	0.013	1.450	0.010	1.680	0.028	1.776	0.036	1.883	0.059	2.005	0.062
$ARMA(1,1)$	1.287	0.089	1.480	0.130	1.734	0.144	1.832	0.142	1.921	0.108	2.026	0.072
$ARMA(2,1)$	1.272	0.034	1.463	0.033	1.702	0.026	1.801	0.021	1.898	0.025	2.017	0.050
$ARMA(1,2)$	1.279	0.027	1.469	0.030	1.718	0.033	1.820	0.027	1.915	0.028	2.024	0.028
$ARMA(3,3)$	1.280	0.023	1.466	0.023	1.705	0.026	1.802	0.029	1.903	0.025	2.025	0.050
$ARMA(4,4)$	1.277	0.028	1.462	0.021	1.698	0.020	1.798	0.022	1.899	0.030	2.020	0.053
$HAR(3)$	1.270	0.024	1.454	0.028	1.690	0.051	1.795	0.059	1.902	0.084	2.025	0.168
$AR(22)$	1.273	0.050	1.459	0.050	1.690	0.035	1.790	0.038	1.893	0.032	2.015	0.030
$AR(30)$	1.280	0.073	1.469	0.062	1.699	0.050	1.795	0.031	1.894	0.030	2.014	0.037
$AR(50)$	1.293	0.136	1.481	0.128	1.716	0.104	1.810	0.098	1.904	0.097	2.018	0.058
$I(1)$	1.406	**0.242**	1.664	**0.238**	2.002	**0.216**	2.169	**0.218**	2.373	**0.211**	2.689	0.190

Table 2.11. Mean of the $RMSE$ and proportion of times the model is in the MCS using the SQ loss function and the T_R statistic at a 95% confidence level.

DGP 3	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.014	0.126	1.086	0.136	1.162	0.125	1.191	0.123	**1.219**	0.104	**1.263**	0.093
$ARFIMA(1,d,0)$	**1.010**	0.023	**1.083**	0.019	**1.160**	0.024	**1.190**	0.018	**1.219**	0.018	**1.263**	0.053
$ARFIMA(0,d,1)$	**1.010**	0.019	**1.083**	0.033	1.161	0.050	1.191	0.075	1.220	0.086	**1.263**	0.136
$ARFIMA(1,d,1)$	1.055	0.014	1.153	0.018	1.310	0.037	1.422	0.060	1.604	0.076	1.566	0.104
$ARFIMA(2,d,1)$	1.061	0.038	1.152	0.046	1.356	0.064	1.493	0.067	1.734	0.076	1.698	0.099
$ARMA(1,1)$	1.015	0.028	1.088	0.037	1.171	0.061	1.204	0.061	1.234	0.054	1.272	0.066
$ARMA(2,1)$	1.015	0.010	1.087	0.014	1.170	0.023	1.203	0.025	1.233	0.027	1.272	0.039
$ARMA(1,2)$	1.015	0.009	1.087	0.012	1.170	0.015	1.203	0.017	1.233	0.021	1.272	0.022
$ARMA(3,3)$	1.017	0.032	1.089	0.045	1.171	0.041	1.203	0.046	1.232	0.043	1.272	0.047
$ARMA(4,4)$	1.016	0.065	1.089	0.056	1.170	0.059	1.203	0.049	1.234	0.041	1.273	0.052
$HAR(3)$	1.011	0.094	1.085	0.091	1.168	0.073	1.201	0.073	1.234	0.106	1.275	**0.186**
$AR(22)$	1.022	0.076	1.095	0.058	1.171	0.050	1.201	0.040	1.230	0.041	1.271	0.038
$AR(30)$	1.024	0.097	1.097	0.085	1.175	0.067	1.203	0.061	1.230	0.068	1.270	0.061
$AR(50)$	1.044	0.154	1.118	0.123	1.192	0.119	1.215	0.102	1.235	0.094	1.271	0.094
$I(1)$	1.286	**0.215**	1.357	**0.228**	1.476	**0.195**	1.534	**0.185**	1.604	**0.162**	1.755	0.151

Table 2.12. Mean of the $RMSE$ and proportion of times the model is in the MCS using the SQ loss function and the T_R statistic at a 95% confidence level.

DGP 4	h=5		10		30		50		100		300	
$d = 0.3$	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS	$RMAD$	MCS
$FI(d)$	1.470	0.128	1.554	0.134	1.597	**0.153**	1.610	**0.191**	1.624	**0.215**	1.637	**0.268**
$ARFIMA(1,d,0)$	1.419	0.043	1.508	0.031	1.565	0.028	1.585	0.031	1.607	0.043	1.626	0.046
$ARFIMA(0,d,1)$	1.471	0.025	1.600	0.024	1.736	0.035	1.805	0.033	1.912	0.039	2.126	0.060
$ARFIMA(1,d,1)$	**1.397**	0.007	**1.495**	0.010	**1.560**	0.017	**1.582**	0.016	**1.605**	0.027	**1.625**	0.033
$ARFIMA(2,d,1)$	1.399	0.015	1.497	0.017	1.561	0.024	**1.582**	0.021	1.606	0.024	**1.625**	0.035
$ARMA(1,1)$	1.420	0.136	1.517	0.124	1.579	0.099	1.597	0.084	1.615	0.080	1.629	0.067
$ARMA(2,1)$	1.414	0.029	1.514	0.033	1.578	0.045	1.596	0.045	1.614	0.045	1.629	0.056
$ARMA(1,2)$	1.405	0.015	1.502	0.020	1.566	0.029	1.587	0.028	1.608	0.030	1.627	0.042
$ARMA(3,3)$	1.404	0.027	1.501	0.029	1.565	0.029	1.585	0.034	1.607	0.043	1.626	0.064
$ARMA(4,4)$	1.406	0.040	1.503	0.041	1.565	0.043	1.584	0.047	1.607	0.053	1.626	0.070
$HAR(3)$	1.423	0.096	1.511	0.105	1.570	0.092	1.590	0.101	1.611	0.109	1.629	0.148
$AR(22)$	1.422	0.041	1.515	0.062	1.572	0.067	1.590	0.061	1.610	0.059	1.627	0.066
$AR(30)$	1.426	0.076	1.520	0.076	1.576	0.086	1.591	0.086	1.611	0.083	1.628	0.077
$AR(50)$	1.442	0.147	1.536	**0.149**	1.595	0.138	1.609	0.117	1.620	0.109	1.630	0.115
$I(1)$	1.786	**0.176**	1.945	0.146	2.072	0.117	2.108	0.105	2.162	0.091	2.315	0.063

D *ARFIMA(1,d,1)*

Table 2.13. Mean of the *RMAD* and proportion of times the model is in the *MCS* using the *MAD* loss function and the T_R statistic at a 95% confidence level.

$\phi_1 = 0.2; \theta_1 = -0.6$	h=5		10		30		50		100		300	
$d = 0.3$	*RMAD*	*MCS*	*RMAD*	*MCS*	*RMAD*	*MCS*	*RMAD*	*MCS*	*RMAD*	*MCS*	*RMAD*	*MCS*
$FI(d)$	0.889	0.078	0.893	0.083	0.897	0.075	0.898	0.065	0.900	0.062	0.902	0.072
$ARFIMA(1,d,0)$	0.886	0.025	0.890	0.027	0.895	0.025	0.897	0.019	0.899	0.021	**0.901**	0.029
$ARFIMA(0,d,1)$	0.885	0.034	0.889	0.033	**0.894**	0.030	**0.896**	0.029	0.899	0.027	**0.901**	0.046
$ARFIMA(1,d,1)$	**0.884**	0.027	**0.888**	0.041	**0.894**	0.057	**0.896**	0.075	**0.898**	0.112	**0.901**	**0.175**
$ARFIMA(2,d,1)$	**0.884**	0.034	0.889	0.035	**0.894**	0.050	**0.896**	0.058	**0.898**	0.076	**0.901**	0.120
$ARMA(1,1)$	0.889	0.054	0.893	0.050	0.897	0.040	0.898	0.041	0.900	0.041	0.902	0.052
$ARMA(2,1)$	0.886	0.029	0.890	0.022	0.895	0.022	0.897	0.025	0.899	0.033	**0.901**	0.045
$ARMA(1,2)$	0.886	0.017	0.890	0.015	0.895	0.019	0.897	0.023	0.899	0.025	**0.901**	0.037
$ARMA(3,3)$	0.886	0.050	0.890	0.056	0.895	0.055	**0.896**	0.049	0.899	0.059	**0.901**	0.078
$ARMA(4,4)$	0.886	0.078	0.890	0.082	0.895	0.098	0.897	0.096	0.899	0.092	**0.901**	0.109
$HAR(3)$	0.885	0.046	0.889	0.058	0.895	0.060	0.897	0.060	0.899	0.070	0.902	0.104
$AR(22)$	0.889	0.093	0.893	0.085	0.896	0.091	0.898	0.086	0.900	0.082	0.902	0.077
$AR(30)$	0.892	0.096	0.896	0.104	0.898	0.114	0.899	0.112	0.900	0.104	0.902	0.100
$AR(50)$	0.896	**0.196**	0.900	**0.186**	0.903	**0.176**	0.903	**0.171**	0.902	**0.165**	0.902	0.157
$I(1)$	1.050	0.144	1.051	0.122	1.057	0.088	1.063	0.089	1.074	0.081	1.106	0.058

E All Competing Models

Figure 2.7. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 1 with different degrees of memory at several horizons.

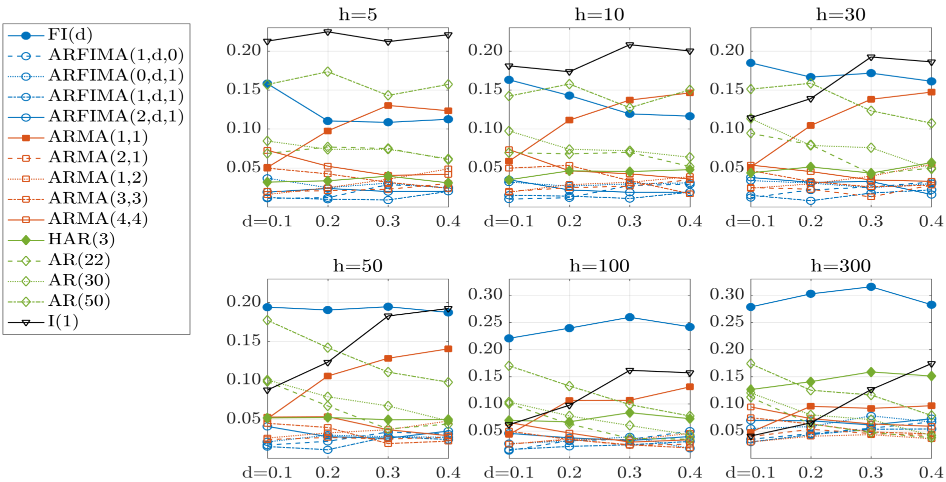


Figure 2.8. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 2 with different theoretical degrees of memory at several horizons.

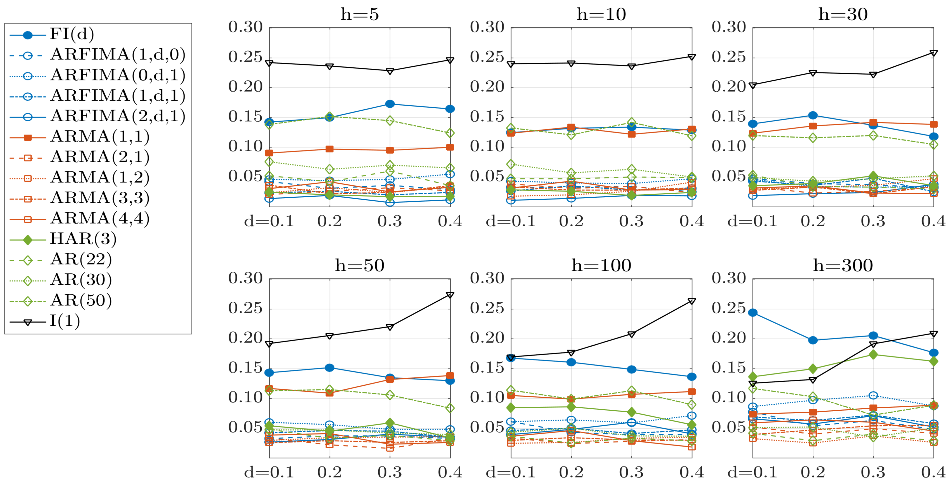


Figure 2.9. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 3 with different degrees of memory at several horizons.

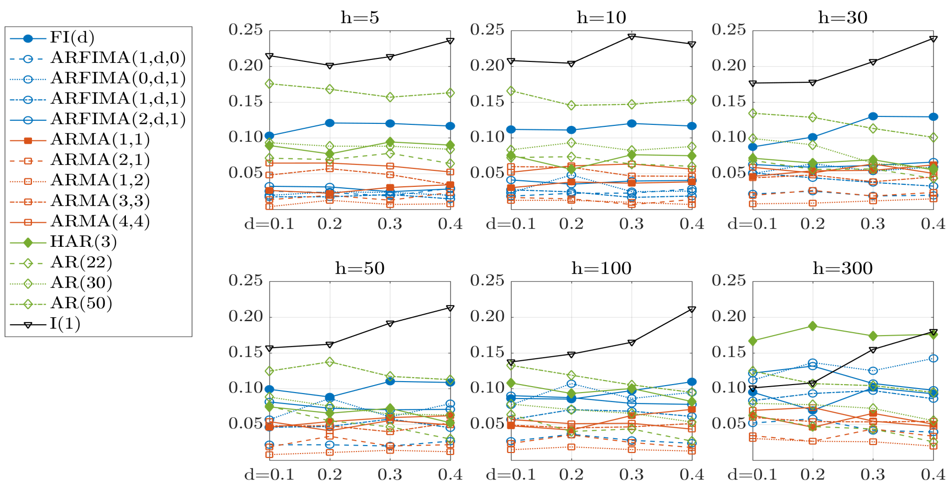
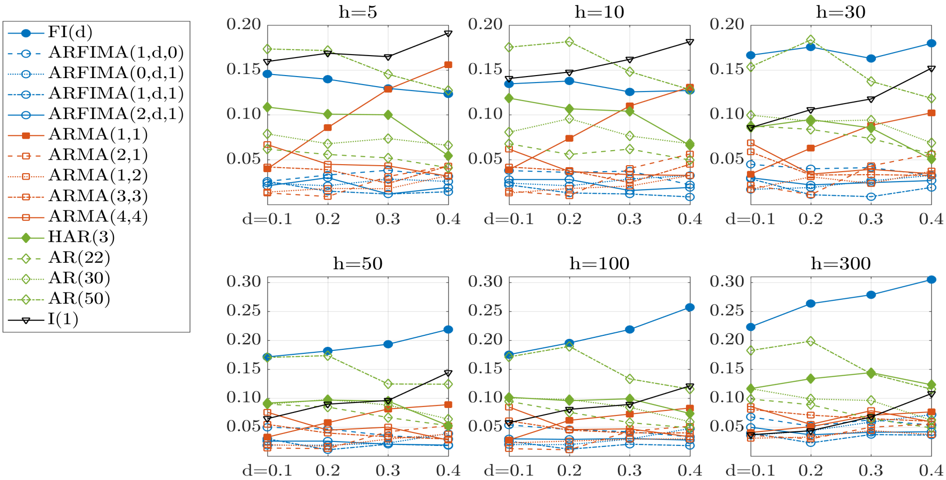


Figure 2.10. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 4 with different degrees of memory at several horizons.



F *HAR* 4

Figure 2.11. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 2 with different degrees of memory at several horizons. For the plots, the starting set contains only the six models shown.

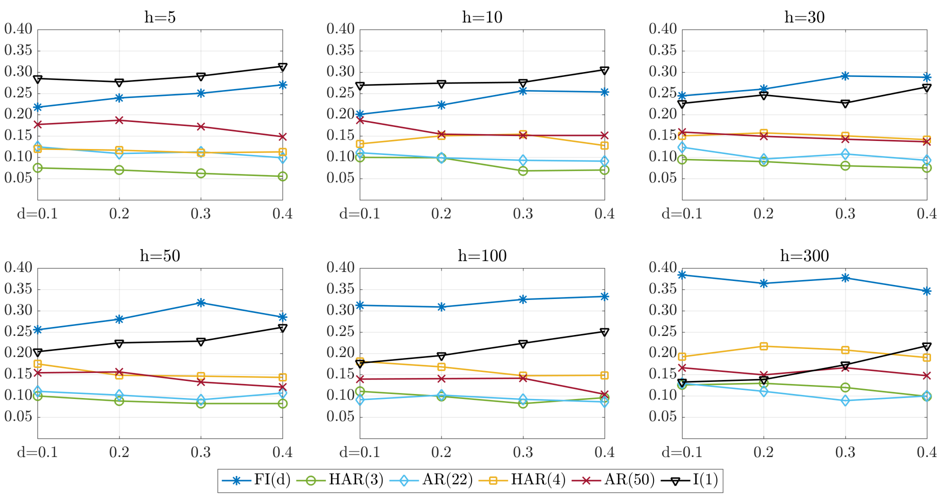


Figure 2.12. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 3 with different degrees of memory at several horizons. For the plots, the starting set contains only the six models shown.

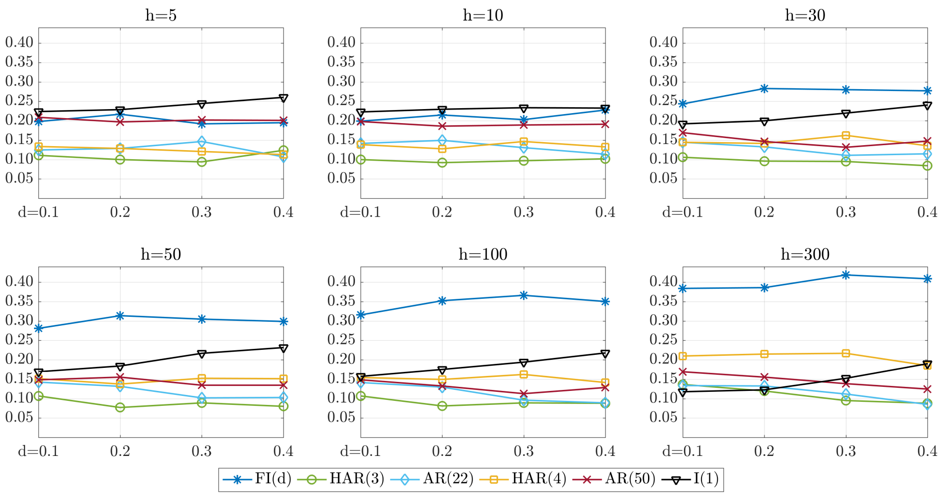
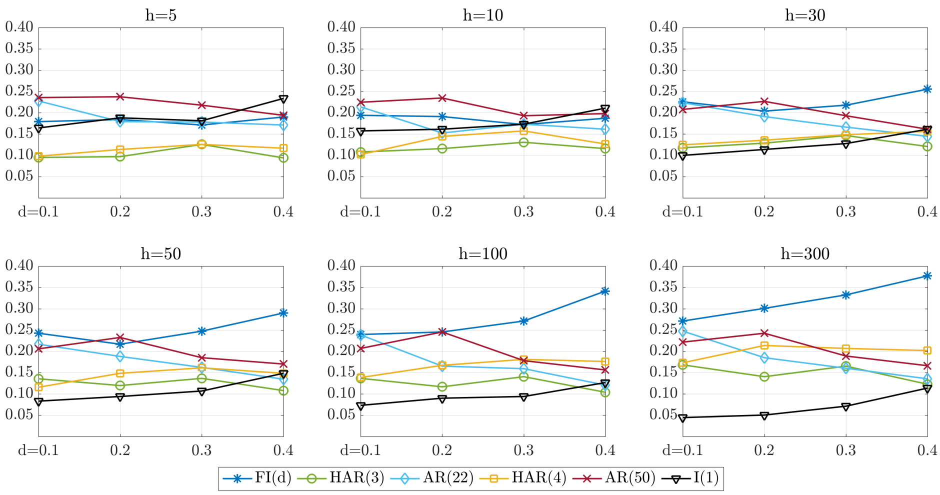


Figure 2.13. Proportion of times the models are in the *MCS* at a 95% confidence level when forecasting *DGP* 4 with different degrees of memory at several horizons. For the plots, the starting set contains only the six models shown.



CHAPTER 3

UNBALANCED REGRESSIONS AND THE PREDICTIVE EQUATION

Daniela Osterrieder
Rutgers Business School and CREATES
Daniel Ventosa-Santaulària
Centro de Investigación y Docencia Económicas
J. Eduardo Vera-Valdés
Aarhus University and CREATES

Abstract

Predictive return regressions with persistent regressors are typically plagued by (asymptotically) biased/inconsistent estimates of the slope, non-standard or potentially even spurious statistical inference, and regression unbalancedness. We alleviate the problem of unbalancedness in the theoretical predictive equation by suggesting a data generating process where returns are generated as linear functions of a lagged latent $I(0)$ risk process. The observed predictor is a function of this latent $I(0)$ process, but it is corrupted by a long memory noise. Such a process may arise due to aggregation or unexpected level shifts. In this setup, the practitioner estimates a misspecified, unbalanced, and endogenous predictive regression. We show that the *OLS* estimate of this regression is inconsistent, but standard inference is possible. To obtain a consistent slope estimate, we then suggest an instrumental variable approach and discuss issues of validity and relevance. Applying the procedure to the prediction of daily returns on the S&P 500, our empirical analysis confirms return predictability and a positive risk-return trade-off.

3.1 Introduction

Returns on financial markets are risky. Investors in financial markets are uncertain about the future value of their investment. Modern portfolio theory (Markowitz, 1952) and the Capital Asset Pricing Model ($CAPM$) of Sharpe (1964) and Lintner (1965) imply that financial market participants care about risk and adjust their return expectations accordingly. Translating the latter statement into a standard dynamic $CAPM$-type argument (see e.g. Glosten, Jagannathan, and Runkle, 1993; Bollerslev, Osterrieder, Sizova, and Tauchen, 2013), expected aggregate market returns, r_t, can be described as

$$\mathrm{E}_t(r_{t+1}) = \gamma\omega_t^2, \tag{3.1}$$

where γ can be thought of as a risk aversion parameter, which according to risk return trade-off theory is expected to be > 0, and ω_t^2 is the local variance of returns with $t = 1, 2, \ldots, T$.

Equation (3.1) implies that given a measure for ω_t^2, returns on the market should be predictable. To investigate the empirical validity of this implication by a statistical linear regression, the researcher needs to identify a proxy for the unobservable local return variance or market risk, ω_t^2. One approach popular in the literature is to find a set of state variables that are assumed to carry information about the unobservable risk, and hence expected returns. Typical predictor variables include the dividend to price ratio (Campbell and Shiller, 1988a; Fama and French, 1988; Cochrane, 1999), the book to market ratio (Lewellen, 1999), the price earnings ratio (Campbell and Shiller, 1988b), interest rate spreads (Fama and French, 1989), and/or the consumption level relative to income and wealth, *cay* (Lettau and Ludvigson, 2001)[1]. A second commonly relied on methodology is to model $\omega_t^2 = \mathrm{Var}_t(r_{t+1})$ explicitly, and estimate its dynamics jointly with the predictive regression within the $(G)ARCH - M$ framework (Engle, Lilien, and Robins, 1987; Engle and Bollerslev, 1986). The recent availability of high-frequency stock market observations has opened a third possibility to proxy for risk, by employing nonparametric techniques to construct realized variance measures (see e.g. Andersen, Bollerslev, Diebold, and Ebens, 2001).

Whichever proxy the researcher decides to chose, they all seem to share the common feature of strong time series persistence. The term spread, measured as the monthly difference between a ten year bond yield and a short-term interest rate by Campbell and Vuolteenaho (2004) and Diebold and Li (2006), has a first-order autocorrelation well above 0.9. The same measure for the price earnings ratio is almost equal to one. Stambaugh (1999) and Lewellen (2004) discover a similarly high

[1] An extensive list of typical predictor variables can be found in Campbell (2000).

correlation estimate for the dividend to price ratio. The latter further reports first-order autocorrelation estimates of 0.99 for the book to market ratio and the earnings price ratio. In the second framework above, the $ARCH$ coefficient or the sum of the $ARCH$ and the $GARCH$ term are typically found to be close to one (for a summary, see e.g. Bollerslev, Chou, and Kroner, 1992). Similarly, the realized variance measures exhibit strong temporal dependence (see e.g. Bollerslev, Tauchen, and Sizova, 2012, and references therein).

The apparent persistence in the proxy for ω_t^2, i.e. the regressor in a predictive return regression, causes econometric problems with estimation and inference that mostly arise due to the correlation between the innovations in the predictor and returns. Firstly, ordinary least squares (OLS) estimation produces a biased and/or inconsistent slope estimate of the predictive regression. If regressors are assumed $I(0)$ with autoregressive dynamics, Stambaugh (1986, 1999) describes the small-sample bias in the OLS estimate. Successively, for instance Kothari and Shanken (1997) and Lewellen (2004) derive estimates that correct for the bias. A large stream of literature describes the regressor dynamics as local to unity (LUR) processes (see e.g. Campbell and Yogo, 2006, and Jansson and Moreira, 2006), thus violating the $I(0)$ assumption. In this setup, the OLS slope estimate has an asymptotic second order bias (Phillips and Lee, 2013). It is not obvious how to correct for the presence of this asymptotic bias since the localizing coefficient cannot be consistently estimated (Phillips, 1987). Torous and Valkanov (2000) further show that if the volatility of the regressor's innovation scaled by the prediction coefficient relative to the volatility of the return innovation decreases sufficiently fast as $T \to \infty$, i.e., at rate T^{-o} with $o > 1$, then the OLS slope estimate of the predictive regression is even inconsistent.

A related econometric problem concerns the statistical inference on the predictability of returns. Within a LUR framework, the t-statistic corresponding to the null hypothesis (H_0) that the regressor contains no predictive information about returns does not converge to the usual normal asymptotic distribution. Similarly, if the regressor instead is assumed to be a fractionally integrated process, $I(d)$, Maynard and Phillips (2001) show that t-statistics have nonstandard limiting distributions. Based on the work of Campbell and Yogo (2006), Cavanagh, Elliott, and Stock (1995), and Stock (1991), who impose the former LUR-type data generating process (DGP) on the regressor, researchers have relied on confidence intervals computed using Bonferroni bounds. Predictability tests relying on this methodology are known to be conservative. A potentially severe drawback of this approach is that the confidence intervals have zero coverage probability if the regressor is stationary, as has been recently shown by Phillips (2014). IVX filtering due to Magdalinos and Phillips (2009) (see also Phillips and Lee, 2013; Gonzalo and Pitarakis, 2012) constitutes an alternative method that resolves the econometric problems of (asymptotic) bias and nonstandard inference in predictive regressions. The underlying idea is to filter the predictor to remove its strong temporal dependence and use the resulting series as an instrument in an instrumental variable (IV) regression. The modified variable addition method of Breitung and Demetrescu (2015), where a redundant regressor is added to the predictive regression, is a further means to achieve standard statistical inference.

A third issue arising in predictive return regressions with persistent regressors that has received less attention is the unbalanced regression phenomenon (see e.g. Banerjee, Dolado, Galbraith, and Hendry, 1993). The studies on predictive regressions with regressor dynamics different from $I(0)$ can be classified into two sets. The first set assumes a DGP where returns are generated as noise, that is under H_0 (see e.g. Maynard and Phillips, 2001). In this setup returns are $I(0)$, whereas regressors are not, making a predictive regression unbalanced *in theory*. The second set of studies (see e.g. Torous and Valkanov (2000)) imposes a return DGP under the alternative hypothesis of predictability (H_1). In this case returns inherit the persistence of the regressor, and hence are not $I(0)$. The predictive regression is balanced *in theory*. Yet, these implications stand in stark contrast to both economic and financial models of expected returns, as well as ample empirical evidence that returns are $I(0)$ processes. It follows that predictive regressions in these frameworks are likely to be unbalanced *in practice*. The alternative DGP of Phillips and Lee (2013) that the authors present in the appendix is one notable exception. Small (or local) deviations from the null hypothesis are explicitly allowed while preserving regression balancedness. Another exception is given in Maynard, Smallwood, and Wohar (2013), who assume a DGP where returns are linearly related to the fractional difference of the regressor rendering returns $I(0)$.

Our work addresses all three econometric issues; that is, bias/consistency, statistical inference, and regression balancedness. We cast our approach in the fractionally integrated modelling framework. There is substantial evidence that observed proxies for risk can be described as $I(d)$ processes, thus possessing long memory. For daily and weekly NASDAQ data on the log price dividend ratio, Cuñado, Gil-Alana, and Perez de Garcia (2005) find an estimate of $d \approx 0.5$. Instead of relying on $(G)ARCH$ models to describe $\omega_t^2 = \text{Var}_t(r_{t+1})$, Baillie, Bollerslev, and Mikkelsen (1996) suggest using a fractionally integrated $GARCH$ ($FIGARCH$) model and find that d is larger than zero but smaller than one for the conditional exchange rate volatility. Similarly, it is well documented that realized variance measures can be modelled as fractionally integrated processes (see, among others, Ding, Granger, and Engle (1993), Baillie et al. (1996), Andersen and Bollerslev (1997), Comte and Renault (1998), Bollerslev et al. (2013)).

Motivated by these empirical regularities, we suggest a *DGP* that linearly relates returns to a latent $I(0)$ predictor, ω_t^2. However, the observed regressor is corrupted by an additive long memory component. Such a *DGP* can be justified by the aggregation idea of Granger (1980) or the presence of structural breaks. Our approach archives balancedness under both hypothesis, the presence as well as the absence of predictability, yet a linear regression of returns on the observed regressor remains unbalanced. We show that in this case the *OLS* estimate in inconsistent, but standard statistical inference based on the t-statistic can be conducted. To cope with the inconsistency, we propose a method that filters the long memory error component without fractional differencing. We prove that the product of a short memory process and a long memory process eliminates the long memory behavior. We then propose to use this device in an IV regression. We prove that the IV estimate is consistent and the corresponding t-statistic is normally distributed. Furthermore, we discuss methods to establish the validity and the relevance of the instruments.

In our empirical application, we demonstrate that our methodology can be used to evaluate intraday return predictability using realized and options-implied variances. We identify two instruments that are closely related to the variance risk premium and the jump component of the stock price process. We find empirical evidence that the latter two are valid and relevant instruments for the options-implied variance of the S&P 500. The IV regression of returns on this proxy for risk results in a positive and significant predictability, providing evidence for a positive risk return trade-off.

3.2 *DGP* and the Unbalanced Predictive Regression

We propose a simple framework that allows for a balanced *DGP* of the prediction target under the null and the alternative hypothesis, while retaining the problem of regression unbalancedness of the type $I(0)/I(d)$ in the empirical prediction model. We assume that the *DGP* of the true predictor variable, x_t^*, is $I(0)$. Throughout the remainder of this work, we assume that x_t^* is unobserved or latent. Further, we assume that there is a function of the true predictor, $x_t = f(x_t^*)$, that is observable. Yet, this variable is corrupted by a fractionally integrated noise, which implies that the observed x_t is $I(d)$. The target, y_t, typically thought of being returns of a risky financial asset, is generated as an $I(0)$ predictive function of x_t^* with prediction coefficient β and level α, such that $\mathrm{E}_t(y_{t+1}) = \alpha + \beta x_t^*$. Equations (3.2)-(3.5) detail the assumed *DGP*.

$$x_t^* = \varepsilon_t \tag{3.2}$$

$$x_t = x_t^* + z_t \tag{3.3}$$

$$y_t = \alpha + \beta x_{t-1}^* + \xi_t \tag{3.4}$$

$$z_t = \left(1-L\right)^{-d}\eta_t, \tag{3.5}$$

where ε_t is independently and identically distributed ($i.i.d.$) with mean zero and variance σ^2_ε, and z_t is stationary fractionally integrated process with $0 < d < 1/2$, such that $(1-L)^d z_t = \eta_t$. L is the usual lag operator and $\eta_t \sim i.i.d.\ (0,\sigma^2_\eta)$. The variance of z_t is $\sigma^2_z = \sigma^2_\eta \frac{\Gamma(1-2d)}{(\Gamma(1-d))^2}$. Finally, $\xi_t \sim i.i.d.\ (0,\sigma^2_\xi)$. Much of the existing work in the field of predictive regressions imposes the assumption that the true predictor, x^*_t, and the observable predictor, x_t, are the same or perfectly correlated. In view of equation (3.1) this would imply that market risk, ω^2_t, were observable. A very different model is considered by Ferson, Sarkissian, and Simin (2003) and Deng (2014). They demonstrate the risk of spurious inference in predictive regressions, where the expected (demeaned) return βx^*_t is assumed to be independent of x_t. Note that both setups can be viewed as extremes of our *DGP*, where the first scenario arises if $\sigma^2_\eta = 0$, and the second scenario occurs if $\beta = 0$ and/or $\sigma^2_\varepsilon = 0$. Instead of imposing these extreme setups, we consider the predictor in our model to be *imperfect*. Similarly to Pastor and Stambaugh (2009) and Binsbergen and Koijen (2010), we assume that the observed variable x_t contains relevant information about the expected return, but it is imperfectly correlated with the latter.

We motivate the assumption that observed regressors are corrupted measures of expected returns by the aggregation result of Granger (1980). Assume that the observed variable x_t in (3.3) is composed of an aggregation of persistent micro units $x_{i,t}$. The predictive regression for returns is typically evaluated for indices; that is, an aggregation of several assets, where the predictor variable would for instance be the dividend to price ratio of an index, the conditional volatility of an index, etc. All of these processes can be viewed as examples of aggregation. Assume that $x_{i,t}$ follows a *DGP* given by

$$x_{i,t} = \phi_i x_{i,t-1} + \vartheta_i w_t + \zeta_{i,t},$$

where w_t and $\zeta_{i,t}$ are independent $\forall\ i$. $\zeta_{i,t}$ are white noise with variance ς^2_i. In addition, there is no feedback in the system, i.e. $x_{i,t}$ does not cause w_t. Thus, $x_{i,t}$ can be viewed as $i = 1,2,\ldots,N$ micro units of a process that are driven by their own past realizations, a common component, w_t, and an idiosyncratic shock, $\zeta_{i,t}$.

Further, assume that the parameters ϕ, ϑ, and ς^2, are drawn from independent populations, and that $\phi \in (0,1)$ is distributed as[2]

$$\mathrm{d}F(\phi^2) = \frac{2}{B(p,l)}\phi^{2p-1}\left(1-\phi^2\right)^{l-1}\mathrm{d}\phi^2 \quad p,l > 1,$$

[2] See Beran, Feng, Ghosh, and Kulik (2013), pp. 85-86.

where $B(\cdot,\cdot)$ denotes the beta function. If we sum the micro units, $x_{i,t}$, we obtain

$$x_t = \sum_{i=1}^{N} \vartheta_i \sum_{j=0}^{\infty} \phi_i^j w_{t-j} + \sum_{i=1}^{N} \sum_{j=0}^{\infty} \phi_i^j \zeta_{i,t-j},$$

where $x_t = \sum_{i=1}^{N} x_{i,t}$. Granger (1980) shows that $x_t \sim I(\delta_x)$, with $\delta_x = \max(1-l+\delta_w, 1-l/2)$, where $w_t \sim I(\delta_w)$. Hence, if we assume that $l = 2(1-d)$ and $\delta_w = 1-2d$, then x_t will be long memory of the order d, i.e. $\delta_x = d \in (0, \frac{1}{2})$. Furthermore, x_t is generated by two components; the first element is a function of the common component w_t, which will be integrated of the order zero. This can be compared to the variable x_t^* in (3.3). The second component is a function of the idiosyncratic error terms $\zeta_{i,t}$, which will be integrated of the order d. This second component can be compared to our variable z_t in the *DGP* of x_t in (3.3). Obviously, in comparison our framework (3.2)-(3.5) is slightly less general, as we make the additional assumption that the innovations of x_t^* and z_t are $i.i.d.$

A different way to motivate our *DGP* for the observable x_t is to think of it as the sum of an expected and an unexpected component. The expected component is correctly centered at the true signal x_t^*. The unexpected component is driven by a process that has (unpredictable) breaks in the level, z_t. The argument that the persistence in observed risk measures may be due to changes in the mean is not new in the literature. For instance, Lettau and van Nieuwerburgh (2008) provide evidence for such structural level changes in the dividend to price ratio, the earning to price ratio, and the book to market ratio. They argue that these patterns could arise as a result of permanent technological innovations that affect the steady-state growth rate of economic fundamentals.

To demonstrate how unexpected structural level breaks can generate long memory dynamics in z_t, we adopt the framework of Diebold and Inoue (2001). Let s_t be a two-state Markov chain, i.e. a random variable that can assume values 1 or 2. s_t is independent of x_t^*. Define

$$\mathscr{P} = \begin{pmatrix} P\{s_t = 1 | s_{t-1} = 1\} & P\{s_t = 1 | s_{t-1} = 2\} \\ P\{s_t = 2 | s_{t-1} = 1\} & P\{s_t = 2 | s_{t-1} = 2\} \end{pmatrix} = \begin{pmatrix} \mathscr{P}_{1,1} & 1-\mathscr{P}_{2,2} \\ 1-\mathscr{P}_{1,1} & \mathscr{P}_{2,2} \end{pmatrix}.$$

Further assume that ϵ_t is a vector of size (2×1), given by

$$\epsilon_t = \begin{cases} (1 \quad 0)' & \text{if } s_t = 1 \\ (0 \quad 1)' & \text{if } s_t = 2 \end{cases}.$$

Now let $z_t = (\varrho_1, \varrho_2)' \epsilon_t$, $\varrho_1 \neq \varrho_2$. That is z_t is a variable that either has level ϱ_1 or ϱ_2, depending on the realization of the Markov chain. We assume that $\mathscr{P}_{1,1} = 1 - c_1 T^{-\delta_1}$,

$\mathscr{P}_{2,2} = 1 - c_2 T^{-\delta_2}$, $\delta_1, \delta_2 > 0$, and $c_1, c_2 \in (0,1)$, and w.l.o.g. that $\delta_1 \geq \delta_2$. If it holds that $\delta_1 < 2\delta_2 < 2 + \delta_1$, then it follows by Diebold and Inoue (2001) that $z_t \sim I(d)$, where $d = \delta_2 - \frac{1}{2}\delta_1$ and $d \in (0, \frac{1}{2})$. In addition, if the parameters satisfy the restriction that $\varrho_1 = -\varrho_2 \frac{c_1}{c_2} T^{\delta_2 - \delta_1}$ then the unconditional mean of z_t, given by[3]

$$\mathrm{E}(z_t) = \frac{\varrho_1 \left(1 - \mathscr{P}_{2,2}\right) + \varrho_2 \left(1 - \mathscr{P}_{1,1}\right)}{2 - \mathscr{P}_{1,1} - \mathscr{P}_{2,2}},$$

is equal to zero. This is in line with our proposed *DGP* of z_t in (3.5). As before, our *DGP* (3.2)-(3.5) is marginally less general. We impose that z_t is a fractional noise, whereas the resulting z_t from the regime switching framework above could have more general $I(d)$ dynamics.

To summarize, our proposed *DGP* (3.2)-(3.5) is consistent with the assumption of *imperfect* predictors. The imperfection is due to an $I(d)$ noise term that corrupts the true signal. This is in line with either viewing the observed predictor as a aggregation of micro units, or assuming that there are unexpected breaks in its level. Our framework further is consistent with the implication of economic/financial models and the empirical evidence that returns are $I(0)$. The *DGP* also incorporates the possibility of return predictability, which is justified by financial models such as (3.1). Finally, our setup allows for strongly persistent observed financial risk factors, which is in line with much of the empirical evidence.

To evaluate the predictability of y_t, the correct regression to estimate would be to regress y_t on x^*_{t-1}. Yet, x^*_{t-1} is not observed by the researcher. We assume that the researcher runs the following misspecified and unbalanced regression

$$y_t = a + b x_{t-1} + e_t. \tag{3.6}$$

This motivates a further feature of our model (3.2)-(3.5). It is a stylized empirical fact that the residuals of (3.6) and the residuals of a time-series model for the predictor are correlated. Consider for instance the regression of stock returns on the dividend to price ratio and an autoregressive model of order one, $AR(1)$. The residuals of the former and the latter typically exhibit a strong negative correlation. Our *DGP* naturally incorporates this property. To see this, we re-write the *DGP* of y_t in (3.4) as

$$y_t = \alpha + \beta x_{t-1} + \left(-\beta z_{t-1} + \xi_t\right). \tag{3.7}$$

Given our *DGP*, it follows that the regression residuals of (3.6) are composed of two elements, that is, $e_t = -\beta z_{t-1} + \xi_t$. Thus, e_t will be naturally correlated with the

[3]See e.g. Hamilton, 1994, p. 684.

innovation in x_t. More precisely, the covariance between the two error terms is given by

$$\mathrm{Cov}(e_t, z_t) = -\beta\sigma_z^2 \frac{d}{1-d}. \tag{3.8}$$

The covariance (3.8) is different from zero, as long as the alternative hypothesis holds, i.e. x_{t-1}^* predicts y_t with $\beta \neq 0$, the long-memory noise term is not constant, i.e. $\sigma_\eta^2 \neq 0$, and $d \in (0, \frac{1}{2})$. This proves that the regression suffers from an endogeneity problem. As we will see, this have repercussions for estimation.

3.3 Ordinary Least Squares Estimation

We describe the implications of regression unbalancedness and endogeneity, where the latter is caused by the correlation between the innovations in the observed noisy regressor and the target, on the *OLS* estimation and inference. Define two matrices $\mathbf{X}$ and $\mathbf{y}$ of size $(T-1) \times 2$ and $(T-1) \times 1$, respectively by

$$\mathbf{X} \equiv \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_{T-1} \end{pmatrix}' \tag{3.9}$$

$$\mathbf{y} \equiv \begin{pmatrix} y_2 & y_3 & \dots & y_T \end{pmatrix}'. \tag{3.10}$$

Theorem 1 summarizes our results for both hypotheses, the presence and absence of return predictability from x_{t-1}^*.

Theorem 1. *Let x_t^*, x_t, and y_t be generated by* (3.2), (3.3), *and* (3.4), *respectively. Estimate regression* (3.6) *by OLS, resulting in*

$$\hat{b}_{OLS} \equiv \left(\hat{a},\, \hat{b}\right)' = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right).$$

Let $\xrightarrow{P}$ denote convergence in probability, and $\xrightarrow{D}$ convergence in distribution. Then, as $T \to \infty$:

1. *If $\beta = 0$*

$$\hat{a} \xrightarrow{P} \alpha \qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_\xi^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)$$

$$T^{-1/2} t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad t_b \xrightarrow{D} \mathcal{N}\left(0, 1\right).$$

$t_a = \hat{a}/\sqrt{\mathrm{Var}(\hat{a})}$ *and* $t_b = \hat{b}/\sqrt{\mathrm{Var}(\hat{b})}$ *denote the* t*-statistics associated with* $\hat{a}$ *and* $\hat{b}$*, respectively, and* $\mathcal{N}(\cdot,\cdot)$ *is the normal distribution. In addition, it holds that* $s^2 \xrightarrow{P} \sigma_\xi^2$*, where* $s^2 = (T-3)^{-1}\sum_{t=2}^{T}\hat{e}_t^2$ *is the variance of the OLS residuals.*

2. *If* $\beta \neq 0$

$$\hat{a} \xrightarrow{P} \alpha \qquad \hat{b} \xrightarrow{P} \beta\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}$$

$$T^{-1/2}t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{1/2}} \qquad T^{-1/2}t_b \xrightarrow{P} \frac{\beta\sigma_\varepsilon^2}{\left(\beta^2\sigma_\varepsilon^2\sigma_z^2+\sigma_\xi^2(\sigma_\varepsilon^2+\sigma_z^2)\right)^{1/2}},$$

where $s^2 \xrightarrow{P} \sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}$.

A proof of Theorem 1 can be found in Appendix A. A crucial result for the proof of the Theorem is the asymptotic distribution of the product of a long memory noise against an $I(0)$ process which we compute in Lemma 1. Once that distribution is obtained, the proof of the Theorem follows a standard procedure finding expressions much in line to the usual ones of *OLS* estimation under short memory measurement error.

The first part of the theorem summarizes the case in which the researcher estimates a predictive regression for unrelated variables in an unbalanced regression framework. In this situation, the *OLS* slope estimate $\hat{b}$ correctly converges to zero and to a normal distribution at the usual rate $T^{-1/2}$. Figure 3.1 compares the empirical distribution of $\hat{b}$ from 200,000 simulations with continuous uniformly distributed errors to the theoretical asymptotic distribution from Theorem 1. Even for small samples of size $T = 250$, the former closely approximates the latter.

In the second part of Theorem 1, we derive the asymptotic inference for the unbalanced regression framework under the alternative hypothesis that there is predictability from x_{t-1}^* on y_t. In this case, *OLS* produces an inconsistent estimate for β. Table 3.3 summarizes the simulated small sample behavior of the relative bias $\hat{b}/\beta$, with errors drawn from t-distributions. These values range from 0.17 to 0.69, which implies a substantial bias towards zero of the *OLS* slope estimate. The table also demonstrates that the bias is not merely present in small samples, as often the relative bias with $T = 1{,}000$ is larger than or equal to the corresponding value with $T = 250$, all else equal. Finally, Table 3.3 shows that $\hat{b}/\beta$ is independent of σ_ξ and β, but it decreases with increasing d and σ_η, and increases with increasing σ_ε. This is fully in line with the theoretical results in Theorem 1. Figure 3.2 plots the empirical average value of $\hat{b}$ for different sample sizes, T, from 200,000 simulations of the *DGP*

(3.2)-(3.5) with t-distributed errors, proving graphical support for the analytical results in the theorem. Taken together, this implies that a non-zero linear relation between the dependent and the independent variable cannot be consistently estimated by *OLS*.

The results reveal that the *OLS* estimate has an asymptotic bias towards zero, which implies that the researcher would underestimate the implied predictive power from x^*_{t-1} on y_t. This finding stands in contrast to the conclusions in Stambaugh (1986, 1999) and Lewellen (2004). Assuming that the covariance between the prediction-regression residuals and the innovations in the predictor is negative, the latter conclude that there is a positive finite-sample bias in the *OLS* prediction estimate stemming from the endogeneity. Hence, if there is positive predictability the researcher will overestimate its magnitude. The problem is somewhat more severe in our setup, as $\hat{b}$ may not merely suffer from a bias, but rather is an inconsistent estimate. Given our assumptions, regression (3.6) is unbalanced in addition to being endogenous. The dependent variable is $I(0)$, whereas the independent variable exhibits long memory, $I(d)$. The *OLS* approach attempts to minimize the sum of squared residuals in the misspecified regression (3.6). This can be achieved by eliminating the memory in e_t, i.e. by letting $\hat{b} \to 0$. This finding is consistent with Maynard and Phillips (2001).

The t-statistic associated with $\hat{b}$ converges asymptotically to a standard normal limiting distribution that is free of nuisance parameters under the null hypothesis that $\beta = 0$. Small sample simulations with t-distributed errors in Table 3.3 support this conclusion. The size of a simple t-test on the parameter is always very close to the nominal size of 5%. Figure 3.1 shows that even for small sample sizes the t-statistic approximates the asymptotic distribution closely. Under the alternative hypothesis, the t-statistic t_b diverges asymptotically at rate $T^{1/2}$. Figure 3.2 supports this conclusion from Theorem 1, plotting the empirical average value of $T^{-1/2}t_b$ for different sample sizes, T, from 200,000 simulations of the *DGP* (3.2)-(3.5) with t-distributed errors. The implication of these results is that one can draw valid statistical inference on the significance of β. A t-test has sufficient asymptotic power to reject the null hypothesis. In other words, with T sufficiently large, the researcher would eventually reject the hypothesis that the parameter is equal to zero. The latter result makes clear that even in the unbalanced and misspecified regression framework considered here, the t-statistic can be considered a useful tool to draw inference on the significance of the predictability of y_t from a latent x^*_{t-1}. Table 3.3 provides small sample simulation evidence for this conclusion. Drawing *DGP* errors from a t-distribution, we find that a t-test generally has good power. Exception from this happen mostly for small sample sizes, $T = 250$, a small absolute value of β, and large d. The worst case scenario occurs when $\sigma_\xi = \sigma_\eta = 1.73$, $\sigma_\varepsilon = 1.13$, $d = 0.49$, and $T = 250$. This is not surprising, as in this case the signal-to-noise ratio of the predictor, $\mathscr{S} \equiv \sigma_\varepsilon / \sigma_z$, is equal to 0.1615, and hence rather small. In addition, the relation between y_t and x^*_{t-1} is blurred by a noise term, ξ_t, that is more volatile than the predictor itself. All else equal, the power increases in $|\beta|$, in σ_ε, and in T; it decreases in d, σ_ξ, and σ_η.

The finding that statistical inference in our unbalanced and endogenous regression framework is not spurious may be somewhat surprising. Generally, these two phenomena when occurring jointly imply a nonstandard limiting distribution of the t-statistic under the null hypothesis. For fractionally integrated regressors, this result can be found in Maynard and Phillips (2001); the case of *LUR* regressors is derived in Cavanagh et al. (1995). Note, however, that given our *DGP* (3.2)-(3.5), the regressor is no longer endogenous under the assumption that $\beta = 0$, that is $\text{Cov}(e_t, z_t) = 0$. From the literature focusing on the traditional $I(0)/I(1)$ unbalanced regression setup with exogenous regressors and $i.i.d.$ innovations, we know that the t-statistic is well behaved and converges to a standard normal random variable, as shown in Noriega and Ventosa-Santaulària (2007) and successively in Stewart (2011). Theorem 1 proves that the same result holds true in our $I(0)/I(d)$ specification.

A further implication of Theorem 1 is that the level of the conditional mean of y_t, α, can be consistently estimated by the *OLS* estimate, $\hat{a}$, independently of the true value of β. Its associated t-statistic, t_a, diverges at rate $T^{1/2}$. Thus, asymptotically the researcher would correctly reject the null hypothesis that $\alpha = 0$ when the null hypothesis is false, based on a simple t-test.

To summarize, the t-statistic corresponding to an *OLS* estimate represents a means to identify the non-existence of a linear relationship between a random variable and its lagged latent predictor. Yet, in the present $I(0)/I(d)$ setup with unobserved regressors, *OLS* yields an inconsistent estimate of such a linear relationship. To cope with the problem of unbalanced regressions, Maynard et al. (2013) suggest to fractionally filter the regressor; fractional differencing has also been applied by Christensen and Nielsen (2007). In this paper, we opt for a different solution to cope with the problem for several reasons. Firstly, the application of the fractional filter to the predictor requires the knowledge of d. As d in not know a priori, the researcher has to estimate it, which introduces an additional degree of uncertainty. Secondly, fractionally differencing the regressor is only a useful approach if the assumed *DGP* for y_t follows:

$$y_t = \alpha + \beta \left(1 - L\right)^d \tilde{x}_{t-1} + \xi_t, \tag{3.11}$$

with $\tilde{x}_t$ being a pure fractionally integrated process. We argue that it is difficult to justify a *DGP* as (3.11) from an economic and financial viewpoint. In a traditional $I(0)/I(1)$ framework, i.e. $d = 1$ in (3.11), the filter $\left(1 - L\right)^d$ applied to $\tilde{x}_{t-1}$ would imply that y_t is driven by the short-run changes of lagged $\tilde{x}_t$, instead of by its level. In the fractionally integrated setup with $d \in (0, \frac{1}{2})$, y_t in (3.11) would be determined by a "hybrid" of levels and changes in the predictor. This is not in line with many economic-financial models. Let y_t be the continuously compounded return on a risky financial asset, or the logarithmic dividend growth; for instance, the Dynamic Gordon Growth Model states that under rational expectations the logarithmic dividend to price ratio in *levels* should have predictive ability for future returns and/or dividend growth. It follows that the true predictor, $\tilde{x}_{t-1}$, cannot be a fractional difference. Similarly, assume y_t is the change in the foreign exchange spot rate and let the predictor be the forward premium; the Forward Rate Unbiasedness theory implies that the expected changed in spot rates is linearly related to the *level* of the forward premium.

Finally, in our assumed DGP, y_t is related to the level of a lagged latent x_t^*, which is corrupted by a persistent error. Fractional differencing in this setup cannot help solving the unbalanced regression problem. Even if d were known, filtering the observed x_{t-1} by $\left(1-L\right)^d$ would imply an over-differencing of the true signal x_{t-1}^*. This would suggest that y_t were driven by an anti-persistent predictor.

3.4 Instrumental Variable Estimation

To alleviate all of the above concerns, we instead propose to estimate the linear relationship by an instrumental variable (IV) approach. Assume that the researcher has access to a valid and relevant $I(0)$ instrument, i.e. a variable that is strongly correlated with x_{t-1}^* but not with the fractional noise, z_{t-1}, and the innovation ξ_t[4]. Theorem 2 summarizes the asymptotic properties of an IV estimation of equation (3.6).

Theorem 2. *Let x_t^*, x_t, and y_t be generated by* (3.2), (3.3), *and* (3.4), *respectively. Assume there exist K instruments*

$$q_{k,t} = \rho_k x_t^* + \upsilon_{k,t}, \quad k = 1, 2, \ldots, K, \tag{3.12}$$

where $\upsilon_{k,t} \sim$ i.i.d. $(0, \sigma^2_{\upsilon_k})$ and $\rho_k \neq 0 \;\forall k$. Further define

$$\mathbf{Q} \equiv \begin{pmatrix} 1 & 1 & \ldots & 1 \\ q_{1,1} & q_{1,2} & \cdots & q_{1,T-1} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,T-1} \\ \vdots & \ddots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,T-1} \end{pmatrix}'. \tag{3.13}$$

[4]Notice that by equation (3.7) it must hold that an instrument that is neither correlated with z_{t-1} nor with ξ_t will by definition also be unrelated to the error term of the unbalanced regression (3.6), e_t.

Estimate regression (3.6) *by* IV *using* $q_{k,t}$ *as instruments for* x_t*. The* IV *estimate is given by*

$$\hat{b}_{IV} \equiv \left(\hat{a},\, \hat{b}\right)' = \left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{y}\right).$$

Then, as $T \to \infty$*:*

1. *If* $\beta = 0$

$$\hat{a} \xrightarrow{P} \alpha \qquad T^{1/2}\hat{b} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_\xi^2\left(\sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2} + 1\right)}{\sigma_\varepsilon^4 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2}}\right)$$

$$T^{-1/2} t_a \xrightarrow{P} \frac{\alpha}{\sigma_\xi} \qquad t_b \xrightarrow{D} \mathcal{N}(0,1),$$

where $s^2 \xrightarrow{P} \sigma_\xi^2$.

2. *If* $\beta \neq 0$

$$\hat{a} \xrightarrow{P} \alpha \qquad \hat{b} \xrightarrow{P} \beta$$

$$T^{-1/2} t_a \xrightarrow{P} \frac{\alpha}{\left(\sigma_\xi^2 + \beta^2\sigma_z^2\right)^{1/2}} \qquad T^{-1/2} t_b \xrightarrow{P} \beta\left(\frac{\sigma_\varepsilon^4 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2}}{\left(\sigma_\xi^2 + \beta^2\sigma_z^2\right)\left(\sigma_\varepsilon^2 \sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2} + 1\right)}\right)^{1/2},$$

where $s^2 \xrightarrow{P} \sigma_\xi^2 + \beta^2\sigma_z^2$.

The proof of Theorem 2 can be found in Appendix A. Analogously to the proof of Theorem 1, a crucial development for the proof of Theorem 2 is the distribution of the product of a long memory process against a short memory one computed in Lemma 1. As the Lemma shows, using $I(0)$ instruments helps in eliminating the long memory behavior of the measurement error, this is the key insight behind the positive results for IV estimation.

Theorem 2 shows that in the absence of predictability, the IV estimate $\hat{b}$ converges to a normal distribution with zero mean at the standard rate $T^{-1/2}$. Figure 3.3 shows that even if $T = 250$, $\hat{b}$ approaches the theoretical asymptotic distribution. More importantly, Theorem 2 demonstrates that IV estimation results in a consistent estimate for β. Hence, under the maintained assumption that the DGP follows (3.2)-(3.5), the predictive power of a latent variable x_{t-1}^* on y_t can be correctly inferred if the researcher finds a relevant and valid instrument for the former. Figure 3.4 supports this conclusion, plotting the average IV estimate $\hat{b}$ over 200,000 simulations for increasing T. The simulation results in Table 3.4 further show that the relative bias, $\hat{b}/\beta$, is very close to one even for small to moderate sample sizes. Across the set of chosen parameter values, the relative bias is bound between 1 and 1.05, when simulated errors are drawn from standard normal distributions. Finally, α can be consistently estimated by IV, as Theorem 2 proves.

Theorem 2 further implies that the statistical significance of β can be correctly inferred from a simple t-test. Under the null hypothesis that $H_0 : \beta = 0$, the t-statistic of the IV estimate $\hat{b}$, t_b, converges to a standard normal distribution, as the simulations in Figure 3.3 confirm. Table 3.4 summarizes the small sample behavior of a t-test. Overall, the size of the test is close to the nominal level of 5%, but on average the test seems somewhat undersized. Nevertheless, the size approaches 5% as T increases, as $|\rho|$ and hence $|\text{Corr}(q_{k,t}, x_t^*)|$ increases, and as σ_ε increases, all else equal. We find the lowest nominal size of approximately 3% for the scenario where $\sigma_\xi = \sigma_\eta = \sigma_\upsilon = 1.73$, $\sigma_\varepsilon = 1.13$, $d = 0.49$, and $T = 250$. As mentioned in Section 3.3, this is to be expected as in this case $\mathscr{S}$ is small.

Contrasting the size of the t-test on the significance of β of the IV estimator in Table 3.4 with the corresponding size for the *OLS* estimator in Table 3.3, we can thus observe that the overall size of the former is smaller than the latter. This does not come as a surprise as the IV estimator is generally less efficient than the *OLS*. Standard errors of the former are comparably slightly larger, leading to a small underrejection of the null hypothesis. It should be noted that, in our setup, there seems to be no risk of detecting predictability too often. This stands in contrast to the usual worry in the literature that predictability tests with persistent regressors may be (heavily) oversized, as pointed out by Elliott and Stock (1994) and Campbell and Yogo (2006) among others, or may even lead to spurious conclusions (Ferson et al., 2003).

Assuming that the variables follow our proposed *DGP* (3.2)-(3.5), we conclude that size is not an issue in our setup. Yet, the power of the *OLS* t-test on the significance of β in the previous section may be insufficient, especially when T is small and d is large. *OLS* hence implies some risk of the researcher not detecting predictability when it is present. Estimation by IV also alleviates this concern. The power of the t-test is very close to 100% across the scenarios that we consider in the simulations in Table 3.4. Asymptotically, t_b diverges at rate $T^{1/2}$ under $H_1 : \beta \neq 0$ as shown in Theorem 2; Figure 3.4 depicts the convergence behavior of the statistic.

3.4.1 Instrument Relevance

As is generally the case, the instrument may not be irrelevant or too weak. To see this, let $q_{k,t} = \upsilon_{k,t}$ in Theorem 2 and estimate regression (3.6) by IV using $q_{k,t}$ as instruments. Then, as $T \to \infty$, $\hat{b} = O_p(1)$.

To demonstrate that choosing an irrelevant instrument can lead to very undesirable properties of the IV estimation, we simulate an instrument as in (3.12) with $\rho_1 = 0$. Table 3.5 shows the the size, power, and relative bias of the resulting IV estimator and the corresponding significance test, when errors are drawn from continuous uniform distributions. The size of a standard t-test on the significance of the prediction coefficient is approximately zero. Similarly, the power of the test is very low, ranging between 0.31% and 14.87%. The lack of power is not a small sample problem, as our simulations show that the power uniformly decreases as T increases, suggesting that asymptotically the probability to reject is zero. This implies that the researcher will tend to conclude that there is no predictability, independent of whether it is present or absent.

The low size and power properties signify that the t-statistic, t_b, is too small in absolute value if the instrument is irrelevant. This may be the result of a too small value of $|\hat{b}|$ and/or of a too large volatility of the estimate, $\sqrt{\text{Var}(\hat{b})}$. Table 3.5 summarizes the relative bias, $\hat{b}/\beta$, which deviates wildly from the reference point of 1. IV estimation with an irrelevant instrument may lead to overestimation, underestimation, or even estimation with the incorrect sign. The relative bias covers a wide range, from -24.52 to 14.72, and the bias is independent of T. Hence, we cannot conclude that $|\hat{b}|$ is too small in absolute value. The low size and power of t-test is therefore mostly a result of a very high variance of the estimator.

To conclude, estimating (3.6) by IV with an irrelevant instrument leads to an inconsistent and inefficient estimator. To avoid such an outcome, we suggest a simple testing procedure. Assume that the researcher has identified a candidate instrument. Recall that the instrument follows the DGP given in (3.12), $q_{k,t} = \rho_k x_t^* + \upsilon_{k,t}$. As x_t^* is unobserved, the researcher cannot simply regress the instrument on x_t^* to conduct inference on the value of ρ_k and thus on the instrument relevance. Instead, $q_{k,t}$ can be regressed on the observed x_t by OLS. By Theorem 1, it holds that the slope coefficient of this regression is an inconsistent estimate of ρ_k, yet valid statistical inference using a t-test can be carried out. Thus, relying on a simple OLS t-test, the researcher can infer whether the instrument is statistically irrelevant.

3.4.2 Instrument Validity

Besides being relevant, the instruments $q_{k,t-1}$ further need to be valid. For an instrument to be valid, it may not be correlated with the residuals of the IV regression of equation (3.6), e_t. To summarize the consequences of IV estimation with an invalid instrument in finite samples, we simulate two types of invalid instruments. The first instrument is correlated with z_{t-1}, i.e., with the fractionally integrated noise. We draw a random series μ_{t-1} from a standard normal distribution with zero mean and variance σ_μ^2, and construct the instrument as in (3.12) with

$$v_{k,t-1} = \mu_{t-1} + \kappa_k z_{t-1}. \tag{3.14}$$

This invalid instrument is correlated with e_t and it is integrated of the order d, $I(d)$. Henceforth, we refer to such instruments as invalid of type 1. Table 3.6 shows size and power properties, as well as the relative bias of an IV estimation of (3.6) with this instrument. The power of a t-test is close to 100% across the considered scenarios. Even when d is large, the researcher will still reject an incorrect null hypothesis in at least 98.60% of the cases. Similarly, the test is correctly sized at 5%. Thus, even when the instrument is invalid and $I(d)$, statistical inference on β using the t-test from the IV estimation can be conducted. Yet, the IV estimate $\hat{b}$ is biased towards zero. As this bias does not disappears with increasing T, we conclude that $\hat{b}$ is inconsistent. The simulated relative bias, $\hat{b}/\beta$, ranges from 0.38 to 0.77. Thus, using such an invalid instrument with innovations given by (3.14) leads to the same outcome as when estimating regression (3.6) by simple OLS.

In practice, it is fairly straightforward for the researcher to avoid invalid instruments of type 1, i.e., that are correlated with z_{t-1} as in (3.14). As they will be integrated of the order d, a simple statistical test for the presence of a fractional root can be relied on. Examples are the Lagrange-Multiplier tests of Robinson (1994) or Tanaka (1999), which test for an integration order d under the null hypothesis against the alternative of an integration order smaller (or larger) than d. The fractional Dickey-Fuller test of Dolado, Gonzalo, and Mayoral (2002) is another possibility that is easy to implement.

We construct a second type of instrument in Table 3.6, which is correlated with ξ_t, and hence correlated with e_t and integrated of the order zero, $I(0)$. We call this second form of instrument invalidity type 2. The innovations of this instrument $q_{k,t-1}$ are simulated as

$$v_{k,t-1} = \mu_{t-1} + \kappa_k \xi_t. \tag{3.15}$$

Table 3.6 shows that choosing such an instrument can have very severe consequences. The size of a t-test on the significance of the prediction coefficient is approximately 100%. The power of the test is also close to 100% in most instances, yet in extreme cases it may drop down to as low as 2.12%. The researcher would therefore be tempted to reach the exact opposite conclusion than what it should be. If there is no predictability, one will always erroneously conclude that there is. Conversely, if there is very strong predictability, i.e., β is bounded far away from zero, we may fail to reject $\beta = 0$. The latter is especially true if d is large, T is small, σ_ξ is small, and σ_η is big.

If the invalid instrument has innovations given by (3.15), the estimation of (3.6) by IV further is strongly inconsistent. The relative bias in Table 3.6, $\hat{b}/\beta$, shows that when $\beta = -2$, $\hat{b}$ is negative but it strongly underestimates the magnitude. When $\beta = 3$, $\hat{b}$ is positive yet it overestimates the magnitude. Finally, when $\beta = 0$,[5] the estimate $\hat{b}$ is positive. We conclude that $\hat{b}$ in this case has a significant positive bias; as it does not decrease as T increases, the estimate is inconsistent.

In practice, using an invalid instrument of type 2 should be avoided at all costs. A common approach to test for the validity of an instrument is to rely on Sargan's $\mathscr{J}$ test (Sargan, 1958). Corollary 1 summarizes the asymptotic behavior of the $\mathscr{J}$ test for our DGP.

Corollary 1. *Let x_t^*, x_t, and y_t be generated by* (3.2), (3.3), *and* (3.4), *respectively. Assume there exist K instruments, generated by* (3.12). *Estimate the following second stage regression by OLS*

$$\hat{\mathbf{e}} = \mathbf{Q}\varpi + \mathbf{v}, \tag{3.16}$$

where $\hat{\mathbf{e}}$ are the regression residuals from regression (3.6) *by IV. ϖ is a $(K+1)$ OLS coefficient vector and $\mathbf{v}$ is a vector of innovations. Compute the uncentered R^2 of regression* (3.16) *as $R_u^2 = 1 - \frac{\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}$. Define a test statistic for the validity of the instruments as*

$$\mathscr{J} \equiv T R_u^2.$$

Then, as $T \to \infty$:

$$\mathscr{J} \xrightarrow{D} \chi^2_{(K-1)}.$$

A proof of Corollary 1 can be found in Appendix A. The corollary shows that even though the true predictor, x_{t-1}^* is not observable, we can still test whether $q_{k,t}$ is a (in)valid instrument of type 2 for the former. The statistical inference on the $\mathscr{J}$-statistic can be based on the standard χ^2 distribution. To evaluate the finite sample properties of the test, we conduct Monte Carlo experiments with 200,000 repetitions. Table 3.7 shows different parameter combinations with $K = 2$, drawing innovations from t-distributions. The simulations are set in a challenging scenario, where we let $\mathrm{Corr}([q_{1,t-1},\ q_{2,t-1}]', x_{t-1}^*) = [0.85,\ 0.1]'$. Thus, there is only one strongly relevant instrument, whereas the second instrument is weakly relevant at best.

The simulation results in Table 3.7 suggest that the $\mathscr{J}$-test is correctly sized at a nominal level of 5%. The test is marginally oversized, with a maximal size of 5.9%

[5] These estimates are not reported here to save space. The results are available from the authors upon request.

across all scenarios, only when $\mathscr{S}$ is small and $\beta \neq 0$. The power of the test is fair for small values of $T = 250$, and is generally good or very good when we let T increase to 1,000. Across all scenarios the power is substantially larger when $\beta = 0$ than when $\beta \neq 0$. This finding is not surprising, as $\hat{e}_t \to \xi_t$ when $\beta = 0$, and hence there is a very clear relation between $q_{k,t-1}$ with innovations as in (3.15) and $\hat{e}_t$. By the same logic, if $\beta \neq 0$ then $\hat{e}_t \to -\beta z_{t-1} + \xi_t$, and hence the signal ξ_t becomes more prominent in $\hat{e}_t$ as d decreases, σ_η decreases, and/or σ_ξ increases, thus increasing the power of the test. Finally, the power of the test decreases as we let $\text{Corr}([q_{1,t-1},\ q_{2,t-1}]', \xi_t)$ decrease from $[0.5,\ -0.6]'$ to $[-0.4,\ 0.3]'$.

We conclude that there is almost no risk of overrejecting instrument validity in finite samples when the instrument is valid of type 2. However, there is a small chance to erroneously conclude that an invalid instrument is valid, due to insufficient power in (very) small samples. To safeguard against this, we recommend that the researcher chose a conservative confidence level, for instance 10%.

3.5 Predicting Returns on the S&P 500

To exemplify that the suggested approach from the previous sections can help alleviate some of the concerns in empirical asset pricing, we predict daily returns, r_{t+1}, $t = 1, 2, \ldots, T$, on the S&P 500 stock market index. We consider the data period from February 2, 2000 until April 25, 2013, resulting in $T = 3325$ observations. We assume that risk or uncertainty in the financial market, i.e. ω_t^2 in (3.1), can be proxied by observable variance measures. Our first risk proxy is the realized return variance, $RV_{\text{RL},t}$, computed on the basis of intradaily observations spaced into 5-minute intervals. Under certain regularity conditions, $RV_{\text{RL},t}$ converges to the daily quadratic variation of returns, as shown by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002). Our second measure is the bipower variation, $BV_{\text{RL},t}$, of Barndorff-Nielsen and Shephard (2004), which converges to the integrated variance of returns. The three series, r_t, $RV_{\text{RL},t}$, and $BV_{\text{RL},t}$, are obtained from the Oxford-Man Institute's "Realised Library"[6]. As a final proxy for ω_t^2, we consider the volatility index, $VIX_{\text{CBOE},t}$. It is a measure for the risk-neutral expectation of return volatility over the next month, and as such can be viewed as the options-implied volatility. We obtain the series $VIX_{\text{CBOE},t}$, which is traded on the Chicago Board of Options Exchange (CBOE), from the WRDS database. We transform the data series into monthly variance units by

$$vix_t^2 = \frac{30}{365} VIX_{\text{CBOE},t}^2.$$

[6] Available at `http://realized.oxford-man.ox.ac.uk/`.

Whereas vix_t^2 is related to the return variation over a month, the raw series $RV_{\mathrm{RL},t}$ and $BV_{\mathrm{RL},t}$ measure daily variation. To align the three measures, we modify the latter two as follows.

$$
\begin{aligned}
rv_t &= \sum_{i=1}^{22} \left(RV_{\mathrm{RL},t-i+1} \times 100^2 + \left\{ \left[\ln \frac{P_{t-i+1}^{(\mathrm{open})}}{P_{t-i}^{(\mathrm{close})}} \right] \times 100 \right\}^2 \right) \\
bv_t &= \sum_{i=1}^{22} \left(BV_{\mathrm{RL},t-i+1} \times 100^2 + \left\{ \left[\ln \frac{P_{t-i+1}^{(\mathrm{open})}}{P_{t-i}^{(\mathrm{close})}} \right] \times 100 \right\}^2 \right).
\end{aligned}
$$

It is well known that the three variance series exhibit strongly dependent dynamics that closely resemble fractionally integrated processes (see e.g. Bollerslev et al., 2013, and references therein). At the same time, asset returns, especially at the daily frequency level, are known to exhibit almost no serial correlation. This renders a regression of the following type unbalanced.

$$r_{t+1} = a + bx_t + e_{t+1}, \tag{3.17}$$

where for the remainder of this section x_t is either rv_t, bv_t, or vix_t^2. To avoid any overlap between daily returns, r_{t+1}, and the realized variance and bipower variation measures, we define returns as intraday net returns[7]

$$r_t = \left[\frac{P_t^{(\mathrm{close})} - P_t^{(\mathrm{open})}}{P_t^{(\mathrm{open})}} \right] \times 100.$$

To provide evidence that regression (3.17) is indeed unbalanced for our data set, we estimate the respective fractional integration order, d_i, of the four series, rv_t, bv_t, vix_t^2, and r_t, jointly.

It is common to rely on semiparametric techniques for the estimation of d_i, as they permit the researcher to assess the long-memory behavior of the process close to frequency zero, while allowing for some unparameterized dynamics at intermediate or high frequencies. There are two commonly used classes of semiparametric estimators; the log-periodogram estimators introduced by Geweke and Porter-Hudak (1983) and the local Whittle estimators, originally developed by Künsch (1987). We rely on the latter class, since it is more robust and efficient, as pointed out by Henry and Zaffaroni (2002). The exact local Whittle (EW) due to Shimotsu and Phillips (2005) is particularly attractive, since it is consistent and asymptotically normally distributed for any value

[7] All estimation results in this section remain virtually unchanged if we rely on daily close-to-close returns, instead. Here we only report the results for intraday returns; outcomes with close-to-close returns are available from the authors upon request.

of d_i. Nielsen and Shimotsu (2007) derive a multivariate version of the EW, which we apply for the joint estimation of d_{rv}, d_{bv}, d_{vix^2}, and d_r[8].

Table 3.8 summarizes our results. The realized variance and the bipower variation are integrated of the order $I(0.35)$ and $I(0.34)$, respectively. At a 5% significance level, we reject that $d_i = 0$ and $d_i = 1$ for both series, yet we fail to reject that $d_i = 0.5$. The point estimate for the memory of the volatility index, vix_t^2, is somewhat higher, $\hat{d}_{vix^2} = 0.44$. However, according to the t-test of Nielsen and Shimotsu (2007) for the equality of d_i, we cannot reject that the three variance series are integrated of the same order. Intraday returns in turn are integrated of the approximate order zero, and we fail to reject $d_i = 0$. The t-tests for $H_0 : d_i = d_j$ indicate the we reject the hypothesis that variance series and returns are integrated of the same order, which makes regression (3.17) unbalanced. For further evidence of the apparently distinct dynamics of the three variance series and stock returns, see also Figure 3.5, where we plot the autocorrelations of the four processes. Whereas shocks to daily returns die out immediately, shocks to rv_t, bv_t, and vix_t^2 are highly persistent. As opposed to the stationary return process, it takes many lags to revert the effect of a shock to the variance.

One shortcoming of the approach above is that the EW is not explicitly robust to the presence of additive perturbations, which are present in three variance processes, rv_t, bv_t, and vix_t^2, under the maintained assumption that x_t follows a short-memory signal plus a long-memory noise process as (3.3). To robustify our approach, we further rely on the trivariate version of the modified EW estimator of Sun and Phillips (2004) (TEW). Let $X_t \equiv [rv_t,\ bv_t,\ vix_t^2]'$. The underlying assumption of the TEW estimation approach in our setup is that the spectral density of X_t at frequency λ is given by

$$f_X(\lambda) \sim D\tau D' + \iota H \quad \text{as } \lambda \to 0+,$$

where $D = (\text{diag}[\lambda^{-d_{rv}},\ \lambda^{-d_{bv}},\ \lambda^{-d_{vix^2}}])$, and τ is a diagonal matrix with elements $f_{\eta_i}(0)$. Hence, we assume that the fractional-noise series, $z_{t,i}$, are uncorrelated across the three variance measures. H is a (3×3) matrix of ones; thus we impose that the signal $x_t^* = \omega_t^2$ is the same for rv_t, bv_t, and vix_t^2, and has variance $2\pi\iota$. We estimate the respective fractional order of integration of the three series jointly with the ratio τ/ι, by concentrated TEW-likelihood. We find that $\hat{d}_{rv}$ =0.36, $\hat{d}_{bv}$ =0.46, $\hat{d}_{vix^2}$ =0.33. The exact asymptotic properties of the TEW are unknown, yet Sun and Phillips (2004) conjecture that the distribution is normal and that standard errors

[8]The consistency and asymptotic properties of the EW estimator rely on the knowledge of the true mean of the data generating process. As this value is not known in practical applications, we modify the EW to account for this uncertainty, relying on the two-step feasible EW estimator of Shimotsu (2010).

are bound between [0.11,0.15]. The estimates for d_i are thus different from zero and statistically indistinguishable from the non-robust estimates in Table 3.8. From the point estimates for τ/ι, we can compute the implied signal-to-noise ratio; we find $\mathscr{S}_{rv}$ =50.12, $\mathscr{S}_{bv}$ =24.28, and $\mathscr{S}_{vix^2}$ =15.44. This suggests that the variation in the signal is strong relative to the volatility in the fractionally integrated noise for all three variance series[9].

Next we investigate the consequences of ignoring the regression unbalancedness and instead estimating the prediction regression by *OLS*. Table 3.9 outlines the results. If we predict daily returns on the S&P 500 by rv_t, the prediction coefficient is very close to zero and it is statistically insignificant. Similarly, if we evaluate the unbalanced regression (3.17) with $x_t = bv_t$, we obtain a very small and insignificant slope estimate. Yet, when we use the vix_t^2 series to predict returns, we find a positive $\hat{b} = 0.15 \times 10^{-2}$ and it is significantly different from zero. The estimated coefficient is very small, however, and we know from Theorem 1 that the estimate is inconsistent.

To alleviate the problems associated with the unbalanced *OLS* regression, we define a set of instruments for *IV* estimation. To that end, note that there is substantial evidence that there is a linear long-run relation between rv_t and vix_t^2 that is $I(0)$. For instance, Bandi and Perron (2006) and Christensen and Nielsen (2006) find evidence of fractional cointegration between the two series. Furthermore, if the cointegrating vector is equal to $[-1,\ 1]'$, then the resulting cointegrating series corresponds to the monthly version of the variance risk premium, vrp_t, as defined by Bollerslev, Tauchen, and Zhou (2009). The latter argue that vrp_t may be viewed as bet on pure volatility; as such it is reasonable to expect that the measure is closely liked to the local variance in (3.1), $\omega_t^2 = x_t^*$, that we are aiming at proxying with the instrument. Bollerslev et al. (2009) and Bollerslev et al. (2013) also present evidence that vrp_t can predict aggregate market returns, which is further motivation for considering the measure to be a relevant instrument in our framework.

Besides the cointegrating relation between rv_t and vix_t^2, we expect that there is a long-run relation between rv_t and bv_t, as both series measure the monthly integrated variance of stock returns. Following the arguments in Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007), and Huang and Tauchen (2005), the cointegrating relation between rv_t and bv_t represents the contribution of price jumps to the variance if the cointegrating vector is equal to $[1,\ -1]'$. For instance, Andersen et al. (2007) find that the jump component exhibits a much lower degree of persistence than the two series rv_t and bv_t, providing evidence for a fractional cointegration

[9] We expect the confidence bands for the estimates for $\mathscr{S}$ to be very wide, and hence their values have to be interpreted with care and rather viewed as indicative. The reason is that the likelihood function for the TEW becomes flat in $(\tau/\iota)^{-1}$ when $T \to \infty$, as shown by Hurvich, Moulines, and Soulier (2005).

relation. Jumps are closely related to stock market volatility; Corsi, Pirino, and Reno (2010) and Andersen et al. (2007), among others, find that the former is an important predictor of the latter. Therefore we anticipate jumps to be a relevant instrument for $\omega_t^2 = x_t^*$.

We investigate the potential cointegration relation by a restricted version of the co-fractional vector autoregressive model of Johansen (2008, 2009) and Johansen and Nielsen (2012), given by

$$\Delta^d X_t = \varphi\left[\theta'\left(1-\Delta^d\right)X_t\right] + \sum_{i=1}^{n}\Gamma_i\Delta^d\left(1-\Delta^d\right)^i X_t + u_t. \tag{3.18}$$

We rely on model (3.18) because it allows us to identify a cointegration relation between the variables, while at the same time explicitly accounting for possible dynamics at higher frequencies, which may be present due to the overlapping nature of rv_t and bv_t[10]. Given the identification problems of the model (see, Carlini and Santucci de Magistris, 2013), we initially fix the cointegration rank $r = 2$ and estimate (3.18) by restricted maximum likelihood. Subsequently, we test for cointegration. For $\hat{d} = 0.38$ (SE($\hat{d}$)=0.10) and $n = 3$ we find the cointegrating matrix estimate

$$\hat{\theta}' = \begin{pmatrix} 1 & -1.1938 & 0 \\ -1.0111 & 0 & 1 \end{pmatrix}.$$

Johansen (2008) states that model (3.18) has a solution and $\theta' X_t \sim I(0)$ if the following conditions are satisfied. Firstly, r needs to be smaller than 3. The value of the likelihood ratio (LR) statistic of Johansen and Nielsen (2012) that provides a test for $H_0 : r \leq 2$ against $r \leq 3$ is equal to 3.7709; thus we fail to reject the null hypothesis. Secondly, it must hold that $|\varphi'_\perp\left(I_{3\times 3} - \sum_{i=1}^n \Gamma_i\right)\theta_\perp| \neq 0$. In our estimation this value is equal to -1.46, i.e. different from zero. Thirdly, the roots c of the characteristic polynomial $|(1-c)I_{3\times 3} - \varphi\theta' c - (1-c)\sum_{i=1}^n \Gamma_i c^i| = 0$ must be either equal to one or $\notin$ a complex disk $\mathbb{C}_d$. Figure 3.6 shows that all roots fulfill this final condition. We conclude that we have identified two instruments

$$q_t = \begin{pmatrix} q_{1,t} \\ q_{2,t} \end{pmatrix} = \hat{\theta}' X_t, \tag{3.19}$$

that are integrated of the order zero. Hence, q_t are not invalid instruments of type 1 as described in Section 3.4.2, that is, q_t is not correlated with the $I(d)$ noise term, z_t.

[10]The Matlab code for the maximum-likelihood estimation of the parameters of model (3.18) has been provided by Nielsen, Popiel, et al. (2014).

If we estimate a restricted version of our benchmark co-fractional model, where $\theta_{(2,1)} = -1$ and $\theta_{(1,2)} = -1$, we obtain a *LR* statistic of 6.6354, which implies that we reject the restriction. Whereas the second cointegrating relation, $q_{2,t}$, is essentially the variance risk premium of Bollerslev et al. (2009), $q_{1,t}$ differs slightly from the pure jump contribution, i.e., the squared jump sizes over one month. More precisely, $q_{1,t} \approx \sum_{i=1}^{22} \sum_{j=1}^{N_{t-i+1}} \psi_{t-i+1,j}^2 - 0.19bv_t$, where $\psi_{t,j}$ is the size of the jth jump on day t, and N_t denotes the total number of jumps in a day. Noting that $|rv_t - bv_t| \geq |q_{1,t}|$ for more than 95% of the total observations in our sample, $q_{1,t}$ thus reduces the absolute value of the jump component by $0.19bv_t$, that is, it sets it closer to zero. This can be viewed as a crude approximation to the standard approach of only considering significant jumps (see, for instance, Tauchen and Zhou, 2011 and Andersen et al., 2007). Relying on the method outlined in Section 3.4.1, we now investigate whether the two instruments are relevant. Regressing $q_{1,t}$ on rv_t, bv_t, and vix_t^2, respectively, we find the corresponding t-statistics, $t_{\hat{\rho}_1}$, to be equal to -6.54, -12.31, and -3.33. The *jump instrument* is a relevant instrument for the unobserved stationary component of all three variance series. Carrying out the same analysis for $q_{2,t}$, we find the respective values for $t_{\hat{\rho}_2}$ to be equal to -11.42, -11.98, 14.11, suggesting that also the *variance risk premium instrument* is strongly relevant.

Table 3.9 lists the outcomes of the *IV* estimations of regression (3.17), using $q_{1,t}$ and $q_{2,t}$ from (3.19) as instruments. If we predict intraday returns with rv_t, we find a negative prediction coefficient, $\hat{b} = -0.013$, that is statistically significant. This finding stands in contrast to the *OLS* estimation result, where we discover that rv_t does not contain an $I(0)$ component that significantly predicts returns. The solution to this puzzle can be found in the the $\mathscr{J}$-test for instrument validity of type 2. The $\mathscr{J}$-statistic is equal to 13.73, which is well above the $\chi^2_{(1)}$ critical value at any commonly considered confidence level. Hence, the *jump instrument* and the *variance risk premium instrument* for the unobserved stationary component in rv_t are invalid. From the simulations in Table 3.6, we know that if the instrument(s) are invalid of type 2, the researcher is likely to find predictability even though there is none. This explains why we erroneously conclude that there is significant return predictability in the series rv_t from the *IV* estimation. Furthermore, the slope estimate $\hat{b}$ is known to have an asymptotic upward bias. Thus, given our results on *OLS* estimation and the fact that the instruments are invalid for *IV* estimation, we find evidence that rv_t does not carry predictive information for daily returns on the S&P 500. Mainly, given that the instruments are invalid we should not trust the *IV* results, while we know that we can draw valid statistical inference on the significance of coefficients estimated by *OLS*. For the bv_t series, the results in Table 3.9 are qualitatively the same.

Finally, we consider vix_t^2 as a predictor. The *OLS* estimation results imply that there is a positive predictability from vix_t^2 on r_{t+1}, but the prediction-coefficient estimate of 0.15×10^{-2} is asymptotically biased towards zero. If we instead predict r_{t+1} by vix_t^2 using the two identified instruments and *IV* estimation, we obtain a statistically significant slope estimate of $\hat{b} = 0.13 \times 10^{-1}$. This estimate is almost nine times larger than the corresponding inconsistent *OLS* estimate. The $\mathscr{J}$-statistic is equal to 1.41. As this value is smaller than the corresponding $\chi^2_{(1)}$ critical value, even if we consider a significance level of 20%, we conclude that *jump* and the *variance risk premium* are valid instruments in this case. Hence, we find strong evidence that there is an unobservable $I(0)$ component, $x_t^* = \omega_t^2$, contained in the vix_t^2 series that positively predicts future daily stock returns, but that is corrupted by a fractionally integrated noise term, z_t. The risk-return trade-off thus is positive.

3.6 Concluding Remarks

This paper presents a novel *DGP* that accounts for many theoretical and empirical features of the return prediction literature, such as persistence in the observed predictors and the stationary noise-type behavior of returns. Assuming that the practitioner estimates a misspecified and unbalanced predictive regression, where the regressors are imperfect measures of the true predictor variable, we show that *OLS* estimation of the predictive regression results in inconsistent estimates for the prediction coefficient. Nevertheless, standard statistical inference based on t-tests remains valid. To avoid the problem of obtaining an inconsistent estimate for the prediction coefficient, we propose a method that filters the long memory error component without fractional differencing. We prove that the product of a short memory process and a long memory process eliminates the long memory behavior. We then propose to use this device in an *IV* regression. If the practitioner has access to a valid and relevant $I(0)$ instrument, *IV* estimation results in a consistent estimate for the predictive coefficient and standard statistical inference on predictability can be carried out.

Our paper is closely related to the work on predictive regressions with *IVX* filtering of Magdalinos and Phillips (2009) and Phillips and Lee (2013), where the predictor is assumed to have *LUR* dynamics. Similarly to our approach, the underlying idea is to find an instrument that is less persistent that the regressor and use it in an *IV* regression. They show that consistency of the prediction estimate and standard statistical inference can be achieved in this framework,[11] which is in line with our

[11] Note that the framework of Phillips and Lee (2013) permits multivariate regressors and discusses multi-period predictions, which we do not consider in this work.

conclusions. From a theoretical viewpoint, Phillips and Lee (2013) also explicitly addresses the issue of an unbalanced regression and their extended framework presented in the appendix permits local deviations from H_0 while retaining balancedness. Important differences to our work are that our setup allows for unrestricted deviations for the null hypothesis of no predictability. Our theoretical predictive equation remains balanced for any value of the prediction coefficient. Secondly, whereas Phillips and Lee (2013) assume that the true predictor is observed, we view regressors as imperfect. Lastly, the instrument in Magdalinos and Phillips (2009) and Phillips and Lee (2013) is easy to find, as it is a filtered version of the predictor itself, and it is relevant and valid by definition. In our setup, the practitioner has to find an instrument and subsequently test for instrument relevance and validity. To that end, we discuss methodologies to investigate instrument relevance and validity.

Finally, we apply the methods outlined in this paper to the investigation of the predictability of daily returns on the S&P 500 stock market. Relying on an analysis of fractional cointegration, we provide one suggestion of how an $I(0)$ instrument can be identified. We find evidence of significant return predictability and a positive risk-return trade-off, using the suggested IV approach.

Acknowledgments

We would like to thank the participants at the 4th Long Memory Symposium, the 8th Nordic Econometric Meeting, and the 2016 Royal Economic Society Conference. Daniela Osterrieder and J. Eduardo Vera-Valdés acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

3.7 References

Andersen, T., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. The Journal of Finance 52 (3), 975–1005.

Andersen, T., Bollerslev, T., Diebold, F., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. The Review of Economics and Statistics 89 (4), 701–720.

Andersen, T., Bollerslev, T., Diebold, F., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 6 (1), 43–76.

Baillie, R. T., Bollerslev, T., Mikkelsen, H. O., 1996. Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 74 (1), 3–30.

Bandi, F. M., Perron, B., 2006. Long memory and the relation between implied and realized volatility. Journal of Financial Econometrics 4 (4), 636–670.

Banerjee, A., Dolado, J., Galbraith, J. W., Hendry, D. F., 1993. Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data (Advanced Texts in Econometrics. Oxford University Press.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Estimating Quadratic Variation Using Realised Variance. Journal of Applied Econometrics 17, 457–477.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Power and Bipower Variation with Stochastic Volatility and Jumps. Journal of Financial Econometrics 2 (1), 1–37.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, USA.

Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes. Springer.

Binsbergen, J. H., Koijen, R. S. J., 2010. Predictive Regressions: A Present-Value Approach. Journal of Finance 65 (4), 1439–1471.

Bollerslev, T., Chou, R. Y., Kroner, K. F., 1992. ARCH Modeling in Finance. Journal of Econometrics 52, 5–59.

Bollerslev, T., Osterrieder, D., Sizova, N., Tauchen, G., 2013. Risk and Return: Long-Run Relations, Fractional Cointegration, and Return Predictability. Journal of Financial Economics 108, 409–424.

Bollerslev, T., Tauchen, G., Sizova, N., 2012. Volatility in Equilibrium: Asymmetries and Dynamic Dependencies. Review of Finance 16 (73), 31–80.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22 (11), 4463–4492.

Breitung, J., Demetrescu, M., 2015. Instrumental Variable and Variable Addition Based Inference in Predictive Regressions. Journal of Econometrics 187 (1), 358–375.

Campbell, J. Y., 2000. Asset pricing at the millennium. The Journal of Finance 55 (4), 1515–1567.

Campbell, J. Y., Shiller, R., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies 1 (3), 195–228.

Campbell, J. Y., Shiller, R., 1988b. Stock prices, earnings, and expected dividends. The Journal of Finance 43 (3), 661–676.

Campbell, J. Y., Vuolteenaho, T., 2004. Bad beta, good beta. The American Economic Review 94 (5), 1249–1275.

Campbell, J. Y., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Carlini, F., Santucci de Magistris, P., 2013. On the Identification of Fractionally Cointegrated VAR Models with the F(d) Condition. CREATES Research Paper 2013-44.

Cavanagh, C., Elliott, G., Stock, J., 1995. Inference in Models with Nearly Integrated Regressors. Econometric Theory 11 (5), 1131–1147.

Christensen, B. J., Nielsen, M. O., 2006. Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. Journal of Econometrics 133, 343–371.

Christensen, B. J., Nielsen, M. O., 2007. The Effect of Long Memory in Volatility on Stock Market Fluctuations. Review of Economics and Statistics 89 (4), 684–700.

Cochrane, J. H., 1999. New facts in finance. Economic Perspectives 23 (3), 36–58.

Comte, F., Renault, E., 1998. Long memory in continuous-time stochastic volatility models. Mathematical Finance 8 (4), 291–323.

Corsi, F., Pirino, D., Reno, R., 2010. Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting. Journal of Econometrics 159 (2), 276–288.

Cuñado, J., Gil-Alana, L. A., Perez de Garcia, F., 2005. A Test for Rational Bubbles in the NASDAQ Stock Index: A Fractionally Integrated Approach. Journal of Banking and Finance 29 (10), 2633–2654.

Deng, A., 2014. Understanding Spurious Regression in Financial Economics. Journal of Financial Econometrics 12 (1), 122–150.

Diebold, F. X., Inoue, A., 2001. Long memory and regime switching. Journal of Econometrics 105 (1), 131–159.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337–364.

Ding, Z., Granger, C. W. J., Engle, R. F., 1993. A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance 1, 83–106.

Dolado, J. J., Gonzalo, J., Mayoral, L., 2002. A Fractional Dickey-Fuller Test for Unit Roots. Econometrica 70 (5), 1963–2006.

Elliott, G., Stock, J. H., 1994. Inference in Time Series Regression when the Order of Integration of a Regressor Is Unknown. Econometric Theory 10, 672–700.

Engle, R. F., Bollerslev, T., 1986. Modelling the Persistence of Conditional Variances. Econometric Reviews 5, 1–50.

Engle, R. F., Lilien, D. M., Robins, R. P., 1987. Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model. Econometrica 55, 391–407.

Fama, E. F., French, K. R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–27.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25 (1), 23–49.

Ferson, W. E., Sarkissian, S., Simin, T. T., 2003. Spurious Regressions in Financial Economics? Journal of Finance 58 (4), 1393–1414.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–237.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48 (5), 1779–1801.

Gonzalo, J., Pitarakis, J., 2012. Regime-Specific Predictability in Predictive Regressions. Journal of Business and Economic Statistics 30 (2), 229–241.

Granger, C. W. J., 1980. Long-Memory Relationships and the Aggregation of Dynamic Models. Journal of Econometrics 14, 227–238.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, Princeton, New Jersey.

Henry, M., Zaffaroni, P., 2002. The long range dependence paradigm for macroeconomics and finance. Columbia University, Department of Economics, Discussion Paper Series.

Huang, X., Tauchen, G., 2005. The Relative Contribution of Jumps to Total Price Variance. Journal of Financial Econometrics 3, 456–499.

Hurvich, C. M., Moulines, E., Soulier, P., 2005. Estimating Long Memory in Volatility. Econometrica 73 (4), 1283–1328.

Jansson, M., Moreira, M., 2006. Optimal Inference in Regression Models with Nearly Integrated Regressors. Econometrica 74 (3), 681–714.

Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. Econometric Theory 24.

Johansen, S., 2009. Representation of cointegrated autoregressive processes with application to fractional processes. Econometric Reviews 28, 121–145.

Johansen, S., Nielsen, M. O., 2012. Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model. Econometrica 80 (6), 2667–2732.

Kothari, S. P., Shanken, J., 1997. Book-to-Market, Dividend Yield, and Expected Market Returns: A Time Series Analysis. Journal of Financial Economics 18, 169–203.

Künsch, H., 1987. Statistical aspects of self-similar processes,. In: Prohorov, Y., Sazonov, V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society. VNU Science Press, pp. 67–74.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. The Journal of Finance 56 (3), 815–849.

Lettau, M., van Nieuwerburgh, S., 2008. Reconciling the Return Predictabililty Evidence. The Review of Financial Studies 21 (4), 1607–1652.

Lewellen, J., 1999. The time-series relations among expected return, risk, and book-to-market. Journal of Financial Economics 54, 5–43.

Lewellen, J., 2004. Predicting Returns with Financial Ratios. Journal of Financial Economics 74, 209–235.

Lintner, J., 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Magdalinos, T., Phillips, P., 2009. Econometric Inference in the Vicinity of Unity. Cowles Foundation Discussion Paper 7, Singapore Management University.

Markowitz, H. M., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

Maynard, A., Phillips, P., 2001. Rethinking an Old Empirical Puzzle: Econometric Evidence on the Forward Discount Anomaly. Journal of Applied Econometrics 16, 671–708.

Maynard, A., Smallwood, A., Wohar, M. E., 2013. Long Memory Regressors and Predictive Testing: A Two-Stage Rebalancing Approach. Econometric Reviews 32 (3), 318–360.

Meddahi, N., 2002. A Theoretical Comparison Between Integrated and Realized Volatility. Journal of Applied Econometrics 17, 479–508.

Nielsen, M., Popiel, M., et al., 2014. A matlab program and user's guide for the fractionally cointegrated var model 1330.

Nielsen, M., Shimotsu, K., 2007. Determining the Cointegrating Rank in Nonstationary Fractional Systems by the Exact Local Whittle Approach. Journal of Econometrics 141 (2), 574–59.

Noriega, A. E., Ventosa-Santaulària, D., 2007. Spurious Regression and Trending Variables. Oxford Bulletin of Economics and Statistics 69 (3), 439–444.

Pastor, L., Stambaugh, R. F., 2009. Predictive Systems: Living with Imperfect Predictors. Journal of Finance 64 (4), 1583–1628.

Phillips, P., 2014. On confidence intervals for autoregressive roots and predictive regression. Econometrica 82 (3), 1177–1195.

Phillips, P. C. B., 1987. Towards a Unified Asymptotic Theory for Autoregression. Biometrika 74 (3), 535–547.

Phillips, P. C. B., Lee, J. H., 2013. Predictive Regression under Various Degrees of Persistence and Robust Long-Horizon Regression. Journal of Econometrics 177, 250–264.

Robinson, P. M., 1994. Efficient Tests of Nonstationary Hypotheses. Journal of the American Statistical Association 89, 1420–1437.

Sargan, D., 1958. The Estimation of Economic Relationships Using Instrumental Variables. Econometrica 26, 393–415.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–443.

Shimotsu, K., 2010. Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend. Econometric Theory 26 (2), 501–540.

Shimotsu, K., Phillips, P. C. B., 2005. Exact local whittle estimation of fractional integration. The Annals of Statistics 33, 1890–1933.

Stambaugh, R. F., 1986. Bias in Regressions with Lagged Stochastic Regressors. University of Chicago, Graduate School of Business, Working Paper Series 156.

Stambaugh, R. F., 1999. Predictive Regressions. Journal of Financial Economics 54 (3), 375–421.

Stewart, C., 2011. A Note on Spurious Significance in Regressions Involving I(0) and I(1) Variables. Empirical Economics 41 (3), 565–571.

Stock, J., 1991. Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series. Journal of Monetary Economics 28 (3), 435–459.

Sun, Y., Phillips, P. C. B., 2004. Understanding the fisher equation. Journal of Applied Econometrics 19 (7), 869–886.

Tanaka, K., 1999. The Nonstationary Fractional Unit Root. Econometric Theory 15, 549–582.

Tauchen, G., Zhou, H., 2011. Realized Jumps on Financial Markets and Predicting Credit Spreads. Journal of Econometrics 160, 102–118.

Torous, W., Valkanov, R., 2000. Boundaries of Predictability: Noisy Preditive Regressions. Working Paper - Anderson Graduate School of Management.

Tsay, W. J., Chung, C. F., 2000. The Spurious Regression of Fractionally Integrated Processes. Journal of Econometrics 96 (1), 155–182.

White, H., 2002. Asymptotic Theory for Econometricians: Revised Edition. Academic Press.

3.8 Appendix

A Proofs

Lemma 1. *Let a_t and b_t be two independent processes given by $a_t = \phi(L)\varepsilon_t$ and $b_t = (1-L)^{-d}\eta_t$ where $\phi(L) = \sum_{i=0}^{\infty}\phi_i L^i$ with $\sum_{i=0}^{\infty} i|\phi_i| < \infty$, $\phi(1) \neq 0$ and $(1-L)^d = \sum_{i=0}^{\infty}\gamma_i L^i$ with $\gamma_i = \Gamma(i+d)/(\Gamma(d)\Gamma(i+1))$, $0 < d < \frac{1}{2}$ and $\epsilon_t \sim i.i.d.(0,\sigma_\varepsilon^2)$, $\eta_t \sim i.i.d.(0,\sigma_\eta^2)$. Define $\mathcal{Z}_t = a_t b_t$; then, $T^{-1/2}\sum_{t=1}^{T}\mathcal{Z}_t/\bar{\sigma} \xrightarrow{D} \mathcal{N}(0,1)$ where $\bar{\sigma}_T^2 := var[T^{-1/2}\sum_{t=1}^{T}\mathcal{Z}_t] \to \bar{\sigma}^2$ as $T \to \infty$.*

Proof:

Let a_t, b_t and $\mathcal{Z}_t$ be as above and let $\mathcal{F}_t$ be the σ-algebra generated by $\{\varepsilon_t, \eta_t, \varepsilon_{t-1}, \eta_{t-1}, \cdots\}$. Note that, given independence, $\mathcal{Z}_t$ is a stationary ergodic process and that $\{\mathcal{Z}_t, \mathcal{F}_t\}$ is an adapted stochastic sequence with $\mathrm{E}[\mathcal{Z}_t^2] = \mathrm{E}[a_t^2 b_t^2] = \sigma_a^2\sigma_b^2 < \infty$ where $\sigma_a^2 = \mathrm{E}[a_t^2]$, $\sigma_b^2 = \mathrm{E}[b_t^2]$.

The lemma follows from Theorem 5.16 in White (2002), where we prove directly that

$$\sum_{m=1}^{\infty}\left(\mathrm{E}\left[\mathrm{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2\right]\right)^{1/2} < \infty.$$

First note that

$$\mathrm{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2 = \mathrm{E}\left[\left(\sum_{i=0}^{\infty}\phi_i\varepsilon_{-i}\right)\left(\sum_{i=0}^{\infty}\gamma_i\eta_{-i}\right)\Bigg|\mathcal{F}_{-m}\right]^2 = \left(\sum_{i=m}^{\infty}\phi_i\varepsilon_{-i}\right)^2\left(\sum_{i=m}^{\infty}\gamma_i\eta_{-i}\right)^2.$$

Thus,

$$\begin{aligned}\sum_{m=1}^{\infty}\left(\mathrm{E}\left[\mathrm{E}[\mathcal{Z}_0|\mathcal{F}_{-m}]^2\right]\right)^{1/2} &= \sum_{m=1}^{\infty}\left(\sigma_\varepsilon^2\sigma_\eta^2\sum_{i=m}^{\infty}\phi_i{}^2\sum_{i=m}^{\infty}\gamma_i{}^2\right)^{1/2} \leq \sum_{m=1}^{\infty}\left(\sigma_\varepsilon^2\sigma_b^2\sum_{i=m}^{\infty}\phi_i{}^2\right)^{1/2} \\ &\leq \sigma_\varepsilon\sigma_b\sum_{m=1}^{\infty}\left(\sum_{i=m}^{\infty}|\phi_i|\right) = \sigma_\varepsilon\sigma_b\left(\sum_{i=0}^{\infty}i|\phi_i|\right) < \infty.\end{aligned}$$

Note in particular that Lemma 1 proves that multiplying the long-memory process by an $I(0)$ process reduces the order of convergence to the one of a short memory process.

Proof of Theorem 1

If $\beta \neq 0$: The *OLS* estimator of regression model (3.6) is given by $\hat{\mathrm{b}}_{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$, where

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \frac{1}{T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}\begin{pmatrix}\sum x_{t-1}^2 & -\sum x_{t-1} \\ -\sum x_{t-1} & T\end{pmatrix},$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \sum y_t \\ \sum y_t x_{t-1} \end{pmatrix},$$

with x_{t-1} and y_t generated by equations (3.3) and (3.4), respectively, and $\hat{\mathfrak{b}}_{OLS} = \left(\hat{a},\, \hat{b}\right)'$. **X** and **y** are defined as in equations (3.9) and (3.10), and all sums run from $t = 1$ to T unless stated otherwise[12]. It follows that

$$\hat{a} = \frac{\sum x_{t-1}^2 \sum y_t - \sum x_{t-1} \sum y_t x_{t-1}}{T \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2} \tag{3.20}$$

$$\hat{b} = \frac{T \sum y_t x_{t-1} - \sum y_t \sum x_{t-1}}{T \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}. \tag{3.21}$$

To derive the asymptotic behavior of the estimators (3.20) and (3.21), along with the associated t-statistics, it is necessary to obtain the limit expression of the sums that appear in the equations. They are summarized in Table 3.1, along with their respective convergence rates. All of the convergence rates (see the *underbraced* expressions) can be found in Tsay and Chung (2000) except for the normalization ratio of $\sum \varepsilon_{t-1} z_{t-1}$ and $\sum \xi_t z_{t-1}$, which follows from Lemma 1.

$\sum x_{t-1}$	=	$\underbrace{\sum \varepsilon_{t-1}}_{O_p(T^{1/2})} + \underbrace{\sum z_{t-1}}_{O_p(T^{d+1/2})}$
$\sum x_{t-1}^2$	=	$\underbrace{\sum \varepsilon_{t-1}^2}_{O_p(T)} + \underbrace{\sum z_{t-1}^2}_{O_p(T)} + 2\underbrace{\sum \varepsilon_{t-1} z_{t-1}}_{O_p(T^{1/2})}$
$\sum y_t$	=	$\alpha T + \beta \sum \varepsilon_{t-1} + \underbrace{\sum \xi_t}_{O_p(T^{1/2})}$
$\sum y_t^2$	=	$\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \underbrace{\sum \xi_t^2}_{O_p(T)} + 2\alpha\beta \sum \varepsilon_{t-1} + 2\alpha \sum \xi_t + 2\beta \underbrace{\sum \xi_t \varepsilon_{t-1}}_{O_p(T^{1/2})}$
$\sum y_t x_{t-1}$	=	$\alpha \sum \varepsilon_{t-1} + \alpha \sum z_{t-1} + \beta \sum \varepsilon_{t-1}^2 + \beta \sum \varepsilon_{t-1} z_{t-1} + \sum \xi_t \varepsilon_{t-1} + \underbrace{\sum \xi_t z_{t-1}}_{O_p(T^{1/2})}$

Table 3.1. Expressions for sums in Theorem 1.

For ease of exposition, denote $\hat{a}^{(n)}$ and $\hat{a}^{(d)}$ the numerator and denominator of $\hat{a}$, respectively, and substitute the expressions from Table 3.1.

$$\hat{a}^{(n)} = \left(\sum x_{t-1}^2\right)\left(\sum y_t\right) - \left(\sum x_{t-1}\right)\left(\sum y_t x_{t-1}\right)$$

[12] Strictly speaking, T should be replaced by $T-1$ in all equations, as we lose one observation by lagging x_t; similarly, all sums should run from $t = 2$ to T. Asymptotically, this will make no difference, however.

$$
\begin{aligned}
&= \underbrace{\alpha T\sum\varepsilon_{t-1}^2+\alpha T\sum z_{t-1}^2}_{O_p(T^2)}-\underbrace{\beta\sum\varepsilon_{t-1}^2\sum z_{t-1}}_{O_p(T^{d+3/2})}-\underbrace{\alpha\left(\sum z_{t-1}\right)^2}_{O_p(T^{2d+1})}\\
&\quad+\underbrace{\sum\xi_t\sum\varepsilon_{t-1}^2+\beta\sum\varepsilon_{t-1}\sum z_{t-1}^2+\sum\xi_t\sum z_{t-1}^2+2\alpha T\sum\varepsilon_{t-1}z_{t-1}}_{O_p(T^{3/2})}\\
&\quad-\underbrace{2\alpha\sum\varepsilon_{t-1}\sum z_{t-1}-\beta\sum z_{t-1}\sum\varepsilon_{t-1}z_{t-1}-\sum z_{t-1}\sum\xi_t\varepsilon_{t-1}-\sum z_{t-1}\sum\xi_t z_{t-1}}_{O_p(T^{d+1})}\\
&\quad+\underbrace{\beta\sum\varepsilon_{t-1}z_{t-1}\sum\varepsilon_{t-1}+2\sum\varepsilon_{t-1}z_{t-1}\sum\xi_t-\sum\varepsilon_{t-1}\sum\xi_t\varepsilon_{t-1}-\sum\varepsilon_{t-1}\sum\xi_t z_{t-1}}_{O_p(T)}\\
&\quad-\underbrace{\alpha\left(\sum\varepsilon_{t-1}\right)^2}_{O_p(T)}
\end{aligned}
$$

$$
\begin{aligned}
\hat{a}^{(d)} &= T\sum x_{t-1}^2-\left(\sum x_{t-1}\right)^2\\
&= \underbrace{T\sum\varepsilon_{t-1}^2+T\sum z_{t-1}^2}_{O_p(T^2)}+\underbrace{2T\sum\varepsilon_{t-1}z_{t-1}}_{O_p(T^{3/2})}-\underbrace{\left(\sum z_{t-1}\right)^2}_{O_p(T^{2d+1})}-\underbrace{2\sum\varepsilon_{t-1}\sum z_{t-1}}_{O_p(T^{d+1})}-\underbrace{\left(\sum\varepsilon_{t-1}\right)^2}_{O_p(T)}.
\end{aligned}
$$

It follows that the expression for $\hat{a}$ simplifies to

$$
\hat{a}=\frac{\alpha T\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{d+3/2})}{T\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{3/2})}.
$$

Dividing both the numerator and the denominator by T^2 and letting $T\to\infty$, we obtain

$$
\operatorname*{plim}_{T\to\infty}\hat{a}=\alpha,
$$

since the remaining terms collapse. Now let $\hat{b}^{(n)}$ and $\hat{b}^{(d)}$ be the numerator and denominator of $\hat{b}$, respectively. Then

$$
\begin{aligned}
\hat{b}^{(n)} &= T\sum y_t x_{t-1}-\left(\sum y_t\right)\left(\sum x_{t-1}\right)\\
&= \underbrace{\beta T\sum\varepsilon_{t-1}^2}_{O_p(T^2)}+\underbrace{\beta T\sum\varepsilon_{t-1}z_{t-1}+T\sum\xi_t\varepsilon_{t-1}+T\sum\xi_t z_{t-1}}_{O_p(T^{3/2})}\\
&\quad-\underbrace{\beta\sum\varepsilon_{t-1}\sum z_{t-1}-\sum\xi_t\sum z_{t-1}}_{O_p(T^{d+1})}-\underbrace{\beta\left(\sum\varepsilon_{t-1}\right)^2-\sum\xi_t\sum\varepsilon_{t-1}}_{O_p(T)}.
\end{aligned}
$$

Noting that $\hat{b}^{(d)}=\hat{a}^{(d)}$, we obtain

$$
\hat{b}=\frac{\beta T\sum\varepsilon_{t-1}^2+O_p(T^{3/2})}{T\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{3/2})}.
$$

Dividing $\hat{b}^{(n)}$ and $\hat{b}^{(d)}$ by T^2, in the limit we have

$$\operatorname*{plim}_{T\to\infty} \hat{b} = \beta \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_z^2}.$$

Next, we demonstrate the derivation of the asymptotic expression for the variance of the regression residuals, s^2.

$$\begin{aligned} s^2 &= \frac{\left(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}_{OLS}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}_{OLS}\right)}{T-2} \\ &= \frac{T\sum y_t^2 \sum x_{t-1}^2 - \sum y_t^2 (\sum x_{t-1})^2 - T\left(\sum y_t x_{t-1}\right)^2 - 2\sum x_{t-1} \sum y_t x_{t-1} \sum y_t + \sum x_{t-1}^2 \left(\sum y_t\right)^2}{(T-2)\left(T\sum x_{t-1}^2 - (\sum x_{t-1})^2\right)}. \end{aligned}$$

Substituting the terms from Table 3.1, we obtain the following expressions for the numerator, $s^{2(n)}$, and denominator, $s^{2(d)}$, of s^2.

$$\begin{aligned} s^{2(n)} &= \underbrace{-T\sum \varepsilon_{t-1}^2 \sum \xi_t^2 - T\sum z_{t-1}^2 \sum \xi_t^2 - T\beta^2 \sum \varepsilon_{t-1}^2 \sum z_{t-1}^2}_{O_p(T^3)} + o_p(T^3) \\ s^{2(d)} &= \underbrace{-T^2 \sum \varepsilon_{t-1}^2 - T^2 \sum z_{t-1}^2}_{O_p(T^3)} + o_p(T^3). \end{aligned}$$

Thus, we obtain the following.

$$s^2 = \frac{\frac{1}{T^2}\left(\sum \xi_t^2 \sum \varepsilon_{t-1}^2 + \sum \xi_t^2 \sum z_{t-1}^2 + \beta^2 \sum z_{t-1}^2 \sum \varepsilon_{t-1}^2\right) + o_p(1)}{\frac{1}{T}\left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) + o_p(1)}.$$

When $T \to \infty$

$$\operatorname*{plim}_{T\to\infty} s^2 = \sigma_\xi^2 + \beta^2 \frac{\sigma_\varepsilon^2 \sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}.$$

Finally, we can write the $t-$statistics as

$$t_a = \hat{a}\left[s^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(1,1)}\right]^{-1/2}$$

$$t_b = \hat{b}\left[s^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(2,2)}\right]^{-1/2}.$$

Following the same procedure as above, we find

$$\begin{aligned} \operatorname*{plim}_{T\to\infty} t_a &= \operatorname*{plim}_{T\to\infty} \hat{a} \times \left(\operatorname*{plim}_{T\to\infty} s^2\right)^{-1/2} \times \operatorname*{plim}_{T\to\infty} \left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(11)}\right)^{-1/2} \\ &= \alpha\left(\sigma_\xi^2 + \beta^2 \frac{\sigma_\varepsilon^2 \sigma_z^2}{\sigma_\varepsilon^2 + \sigma_z^2}\right)^{-1/2} \left(\operatorname*{plim}_{T\to\infty} \frac{\sum x_{t-1}^2}{T\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}\right)^{-1/2} \end{aligned}$$

$$= \alpha\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\frac{\operatorname*{plim}_{T\to\infty}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2+O_p(T^{1/2})\right)}{\operatorname*{plim}_{T\to\infty}\left(T\sum\varepsilon_{t-1}^2+T\sum z_{t-1}^2+O_p(T^{3/2})\right)}\right)^{-1/2}$$

$$= \alpha\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\frac{\sigma_\varepsilon^2+\sigma_z^2}{\operatorname*{plim}_{T\to\infty}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)}\right)^{-1/2}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}t_a = \alpha\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}, \tag{3.22}$$

and

$$\operatorname*{plim}_{T\to\infty} t_b = \operatorname*{plim}_{T\to\infty}\hat{b}\times\left(\operatorname*{plim}_{T\to\infty}s^2\right)^{-1/2}\times\operatorname*{plim}_{T\to\infty}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(22)}\right)^{-1/2}$$

$$= \left(\beta\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T}{T\sum x_{t-1}^2-\left(\sum x_{t-1}\right)^2}\right)^{-1/2}$$

$$= \left(\beta\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)\left(\sigma_\xi^2+\beta^2\frac{\sigma_\varepsilon^2\sigma_z^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}\left(\frac{1}{\operatorname*{plim}_{T\to\infty}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2+O_p(T^{1/2})\right)}\right)^{-1/2}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}t_b = \left(\frac{\beta^2\sigma_\varepsilon^4}{\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)+\beta^2\sigma_\varepsilon^2\sigma_z^2}\right)^{1/2}. \tag{3.23}$$

If $\beta=0$: Note that the asymptotic behavior for $\hat{a}$ does not change since the terms with the largest order of divergence in $\hat{a}^{(d)}$ and $\hat{a}^{(d)}$ do not involve β. Similarly, the asymptotic behavior of $\hat{b}^{(d)}$ remains the same, yet the limit of $\hat{b}^{(n)}$ is different if $\beta=0$. We find that

$$\hat{b} = \frac{T\left(\sum\xi_t\varepsilon_{t-1}+\sum\xi_t z_{t-1}\right)+o_p(T^{3/2})}{T\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{3/2})}$$

$$= \frac{\frac{1}{T^{1/2}}\frac{1}{T^{1/2}}\left(\sum\xi_t\varepsilon_{t-1}+\sum\xi_t z_{t-1}\right)+o_p(T^{-1/2})}{\frac{1}{T}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{-1/2})}$$

$$T^{1/2}\hat{b} = \frac{\frac{1}{T^{1/2}}\left(\sum\xi_t\left(\varepsilon_{t-1}+z_{t-1}\right)\right)+o_p(1)}{\frac{1}{T}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+O_p(T^{-1/2})}. \tag{3.24}$$

The denominator of (3.24) converges in probability to $\sigma_\varepsilon^2+\sigma_z^2$. The numerator involves the sum of the random variable $\xi_t\left(\varepsilon_{t-1}+z_{t-1}\right)$, which from the Lemma has constant variance $\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)$. Thus, the numerator converges in distribution to $\mathcal{N}\left(0,\sigma_\xi^2\left(\sigma_\varepsilon^2+\sigma_z^2\right)\right)$. It follows that

$$T^{1/2}\hat{b}\xrightarrow{D}\mathcal{N}\left(0,\frac{\sigma_\xi^2}{\sigma_\varepsilon^2+\sigma_z^2}\right). \tag{3.25}$$

The asymptotic behavior of the numerator of s^2, i.e. $s^{2(n)}$ changes if $\beta=0$, whereas for $s^{2(d)}$ there is no change. As a result, we get

$$\operatorname*{plim}_{T\to\infty} s^2=\operatorname*{plim}_{T\to\infty}\frac{\frac{1}{T^2}\left(\sum\xi_t^2\sum\varepsilon_{t-1}^2+\sum\xi_t^2\sum z_{t-1}^2\right)+o_p(1)}{\frac{1}{T}\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2\right)+o_p(1)}=\sigma_\xi^2.$$

Plugging this result into the expression for t_a in (3.22), it follows that $\operatorname*{plim}_{T\to\infty} T^{-1/2}t_a=\frac{\alpha}{\sigma_\xi}$. Finally, we find the asymptotic behavior of t_b in the case where $\beta=0$. We re-write (3.23) as

$$\begin{aligned} t_b &= \hat{b}\times\left(s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(22)}\right)^{-1/2} \\ &= \hat{b}\times\left(s^2\frac{1}{\left(\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2+O_p(T^{1/2})\right)}\right)^{-1/2} \\ &= T^{1/2}\hat{b}\times\left(s^2\frac{1}{\left(\frac{1}{T}\sum\varepsilon_{t-1}^2+\sum z_{t-1}^2+O_p(T^{-1/2})\right)}\right)^{-1/2}. \end{aligned} \tag{3.26}$$

The first term in (3.26), $T^{1/2}\hat{b}$ converges in distribution to a normal by (3.25). The second term converges in probability to $\left(\frac{\sigma_\xi^2}{\sigma_\varepsilon^2+\sigma_z^2}\right)^{-1/2}$. Hence,

$$t_b\xrightarrow{D}\mathcal{N}(0,1).$$

Proof of Theorem 2

This section presents proofs for the asymptotic results in Theorem 2. The IV estimator of regression model (3.6) is given by

$$\hat{\mathfrak{b}}_{IV}\equiv\left(\hat{a},\,\hat{b}\right)'=\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{y}\right),$$

where $\mathbf{Q}$, $\mathbf{X}'$, and $\mathbf{y}$ are defined in equations (3.13), (3.9), and (3.10), respectively. Introduce the following auxiliary notation

$$\mathbf{Q}'\mathbf{X} = \begin{pmatrix} T & \sum x_{t-1} \\ \sum q_{1,t-1} & \sum x_{t-1}q_{1,t-1} \\ \vdots & \vdots \\ \sum q_{K,t-1} & \sum x_{t-1}q_{K,t-1} \end{pmatrix} \equiv \begin{pmatrix} \underset{1\times 1}{T} & \underset{1\times 1}{\chi} \\ \underset{K\times 1}{\mathbf{q}} & \underset{K\times 1}{\mathbf{r}} \end{pmatrix} \tag{3.27}$$

$$\mathbf{Q}'\mathbf{y} = \begin{pmatrix} \sum y_t \\ \sum y_t q_{1,t-1} \\ \vdots \\ \sum y_t q_{K,t-1} \end{pmatrix} \equiv \begin{pmatrix} \underset{1\times 1}{y} \\ \underset{K\times 1}{\mathbf{t}} \end{pmatrix} \tag{3.28}$$

$$\mathbf{Q}'\mathbf{Q} = \begin{pmatrix} T & \sum q_{1,t-1} & \dots & \sum q_{K,t-1} \\ \sum q_{1,t-1} & \sum q_{1,t-1}^2 & \dots & \sum q_{1,t-1}q_{K,t-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum q_{K,t-1} & \sum q_{K,t-1}q_{1,t-1} & \dots & \sum q_{K,t-1}^2 \end{pmatrix}$$

$$\equiv \begin{pmatrix} \underset{1\times 1}{T} & \underset{1\times K}{\mathbf{q}'} \\ \underset{K\times 1}{\mathbf{q}} & \underset{K\times K}{\mathbf{B}} \end{pmatrix} \tag{3.29}$$

It follows that

$$(\mathbf{Q}'\mathbf{Q})^{-1} = \frac{1}{c}\begin{pmatrix} 1 & -\mathbf{q}'\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{q} & c\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1} \end{pmatrix},$$

where $\underset{1\times 1}{c} \equiv T - \mathbf{q}'\mathbf{B}^{-1}\mathbf{q}$. Furthermore,

$$\mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1} = \frac{1}{c}\begin{pmatrix} c & 0 \\ \chi - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q} & -\chi\mathbf{q}'\mathbf{B}^{-1} + c\mathbf{r}'\mathbf{B}^{-1} + \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1} \end{pmatrix},$$

and

$$\begin{aligned} \mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X} &= \frac{1}{c}\begin{pmatrix} Tc & \chi c \\ \chi c & \left(\chi - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(\chi - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} \end{pmatrix} \\ \mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{y} &= \frac{1}{c}\begin{pmatrix} yc \\ y\chi - y\mathbf{r}'\mathbf{B}^{-1}\mathbf{q} + \left(-\chi\mathbf{q}'\mathbf{B}^{-1} + c\mathbf{r}'\mathbf{B}^{-1} + \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\mathbf{B}^{-1}\right)\mathbf{t} \end{pmatrix}. \end{aligned}$$

Now, note that the following relation must hold true

$$\left(\mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1} = \frac{c}{Tc\left(\chi - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(\chi - \mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)' + Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - \chi^2c^2} \times$$

$$\left(\begin{array}{cc} \left(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)'+c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} & -\chi c \\ -\chi c & Tc \end{array}\right).$$

Thus, the IV estimate can be re-written as follows.

$$\hat{\mathsf{b}}_{IV} = \frac{1}{Tc\left(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)\left(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\right)'+Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2c^2}\times$$
$$\left(\begin{array}{c} \left(\chi\mathbf{q}'-c\mathbf{r}'-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\right)c\mathbf{B}^{-1}\left(-y\mathbf{r}+\chi\mathbf{t}\right) \\ -\left(\chi\mathbf{q}'-c\mathbf{r}'-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q}\mathbf{q}'\right)Tc\mathbf{B}^{-1}\mathbf{t}+\left(\chi\mathbf{q}'-T\mathbf{r}'\right)yc\mathbf{B}^{-1}\mathbf{q} \end{array}\right).$$

As for the proof of Theorem 1 in Section A, it is necessary to obtain the limit expression of the sums that appear in the definitions of the IV estimates and the associated t-ratios. Most of these expressions are summarized in Table 3.1. The remaining sums can be found in Table 3.2.

$\sum q_{k,t-1}$	=	$\rho_k\sum\varepsilon_{t-1}+\underbrace{\sum v_{k,t-1}}_{O_p(T^{1/2})}$
$\sum q_{k,t-1}^2$	=	$\underbrace{\rho_k^2\sum\varepsilon_{t-1}^2}_{O_p(T)}+\underbrace{\sum v_{k,t-1}^2}_{O_p(T)}+2\underbrace{\rho_k\sum\varepsilon_{t-1}v_{k,t-1}}_{O_p(T^{1/2})}$
$\sum y_t q_{k,t-1}$	=	$\alpha\rho_k\sum\varepsilon_{t-1}+\alpha\sum v_{k,t-1}+\beta\rho_k\sum\varepsilon_{t-1}^2+\beta\sum\varepsilon_{t-1}v_{k,t-1}+\rho_k\underbrace{\sum\xi_t\varepsilon_{t-1}}_{O_p(T^{1/2})}+\underbrace{\sum\xi_t v_{k,t-1}}_{O_p(T^{1/2})}$
$\sum x_{t-1}q_{k,t-1}$	=	$\rho_k\sum\varepsilon_{t-1}^2+\sum\varepsilon_{t-1}v_{k,t-1}+\rho_k\sum\varepsilon_{t-1}z_{t-1}+\underbrace{\sum z_{t-1}v_{k,t-1}}_{O_p(T^{1/2})}$
$\sum q_{k,t-1}q_{j,t-1}$	=	$\rho_k\rho_j\sum\varepsilon_{t-1}^2+\rho_k\sum\varepsilon_{t-1}v_{j,t-1}+\rho_j\sum\varepsilon_{t-1}v_{k,t-1}+\underbrace{\sum v_{k,t-1}v_{j,t-1}}_{O_p(T^{1/2})}$

Table 3.2. Expressions for sums in Theorem 2 with $j\neq k$; $k=1,\cdots,K$.

From the expressions in Tables 3.1 and 3.2 it follows that all elements of $\mathbf{B}$ are of the order $O_p(T)$. Hence, it must hold that $\mathbf{B}^{-1}$ is of order $O_p(T^{-1})$. Furthermore, the elements of all vectors have the same convergence rates; i.e. $\mathbf{q}$ is $O_p(T^{1/2})$, $\mathbf{r}$ is $O_p(T)$, and $\mathbf{t}$ is $O_p(T)$ if $\beta\neq 0$ and $O_p(T^{1/2})$ otherwise. Finally, we note the following orders for the scalars. χ is $O_p(T^{d+1/2})$, y is $O_p(T)$, and c is $O_p(T)$.

Let $\hat{a}^{(n)}$ denote the numerator of $\hat{a}$, given by (3.30). Note that independently of the true value of β, we find the following dominant terms

$$\hat{a}^{(n)} = \underbrace{c^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}_{O_p(T^4)} + o_p(T^4).$$

Similarly, the denominator of $\hat{a}$ given in (3.30) is

$$\hat{a}^{(d)} = \underbrace{T c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}_{O_p(T^4)} + o_p(T^4).$$

It follows that in the limit we can write

$$\operatorname*{plim}_{T\to\infty} \hat{a} = \operatorname*{plim}_{T\to\infty} \frac{\frac{1}{T^4} c^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + o_p(1)}{\frac{1}{T^3} c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + o_p(1)} = \operatorname*{plim}_{T\to\infty} \frac{y}{T} = \frac{\alpha T}{T} = \alpha \tag{3.30}$$

Now, note that the denominator of $\hat{b}$ is identical to the denominator of $\hat{a}$. Hence $\hat{b}^{(d)} = \hat{a}^{(d)}$. For the limiting behavior of the numerator of $\hat{b}$, we need to distinguish between $\beta = 0$ and $\beta \neq 0$. First, define additional auxiliary variables. Let

$$\mathbf{p} \equiv \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_K \end{pmatrix} \quad \mathbf{u} \equiv \begin{pmatrix} \sum v_{1,t-1} \\ \vdots \\ \sum v_{K,t-1} \end{pmatrix} \quad \mathbf{v} \equiv \begin{pmatrix} \sum v_{1,t-1}^2 \\ \vdots \\ \sum v_{K,t-1}^2 \end{pmatrix} \quad \mathbf{w} \equiv \begin{pmatrix} \sum \xi_t v_{1,t-1} \\ \vdots \\ \sum \xi_t v_{K,t-1} \end{pmatrix}.$$

If $\beta \neq 0$: Following the convergence rates in the tables, we find

$$\hat{b}^{(n)} = \underbrace{T c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{t}}_{O_p(T^4)} + O_p(T^{7/2}).$$

Hence,

$$\operatorname*{plim}_{T\to\infty} \hat{b} = \operatorname*{plim}_{T\to\infty} \frac{\frac{1}{T^3} c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{t} + O_p(T^{-1/2})}{\frac{1}{T^3} c^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + o_p(1)} = \operatorname*{plim}_{T\to\infty} \frac{\mathbf{r}'\mathbf{B}^{-1}\mathbf{t}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}} = \operatorname*{plim}_{T\to\infty} \frac{\mathbf{r}'\mathbf{B}^{-1}\beta\sum\varepsilon_{t-1}^2\mathbf{p}}{\mathbf{r}'\mathbf{B}^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}} = \beta$$

Next we investigate the asymptotic behavior of s^2, which is defined as

$$s^2 = \frac{\left(\mathbf{y} - \mathbf{X}\hat{\mathfrak{b}}_{IV}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\mathfrak{b}}_{IV}\right)}{T-2}.$$

Introduce the additional auxiliary notation $h \equiv \sum y_t^2$, $o \equiv \sum x_{t-1} y_t$, and $m \equiv \sum x_t^2$. The respective orders are $O_p(T)$, $O_p(T)$ if $\beta \neq 0$ and $O_p(T^{d+1/2})$ otherwise, and $O_p(T)$. We can re-write s^2 as

$$s^2 = \frac{h\hat{b}^{(d)} - 2y\hat{a}^{(n)} - 2o\hat{b}^{(n)} + T\hat{a}\hat{a}^{(n)} + 2\chi\hat{a}^{(n)}\hat{b} + m\hat{b}\hat{b}^{(n)}}{(T-2)\hat{b}^{(d)}}.$$

Note that the denominator of s^2, $s^{2(d)}$, is equal to $(T-2)a^{(d)} = (T-2)b^{(d)}$. For the case where $\beta \neq 0$, we can write the numerator of s^2, $s^{2(n)}$, in the limit as

$$s^{2(n)} = \underbrace{\hbar \hat{b}^{(d)} - 2y\hat{a}^{(n)} - 2o\hat{b}^{(n)} + T\hat{a}\hat{a}^{(n)} + m\hat{b}\hat{b}^{(n)}}_{O_p(T^5)} + o_p(T^5),$$

and hence by plugging in the dominant terms, we find

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} s^{2(n)} &= \operatorname*{plim}_{T\to\infty} \Big\{ \left(\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \sum \xi_t^2\right) Tc^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - 2\alpha Tc^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} \\
&\quad -2\beta \sum \varepsilon_{t-1}^2 Tc^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{t} + T\alpha c^2 y \mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + \left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) \beta Tc^2 \mathbf{r}'\mathbf{B}^{-1}\mathbf{t} \\
&\quad + o_p(T^5) \Big\} \\
&= \operatorname*{plim}_{T\to\infty} \Big\{ \left(\alpha^2 T + \beta^2 \sum \varepsilon_{t-1}^2 + \sum \xi_t^2\right) T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 \\
&\quad -2\alpha^2 T^4 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 - 2\beta^2 \left(\sum \varepsilon_{t-1}^2\right)^3 T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \\
&\quad + T^4 \alpha^2 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 \\
&\quad + \left(\sum \varepsilon_{t-1}^2 + \sum z_{t-1}^2\right) \beta^2 T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 + o_p(T^5) \Big\} \\
&= \operatorname*{plim}_{T\to\infty} \Big\{ T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 \left(\sum \xi_t^2 + \beta^2 \sum z_t^2\right) + o_p(T^5) \Big\}
\end{aligned}$$

We can therefore conclude that

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} s^2 &= \operatorname*{plim}_{T\to\infty} \frac{T^3 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 \left(\sum \xi_t^2 + \beta^2 \sum z_t^2\right) + o_p(T^5)}{T^4 \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 + o_p(T^5)} \\
&= \operatorname*{plim}_{T\to\infty} \frac{T^{-2} \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2 \left(\sum \xi_t^2 + \beta^2 \sum z_t^2\right)}{T^{-1} \mathbf{p}'\mathbf{B}^{-1}\mathbf{p} \left(\sum \varepsilon_{t-1}^2\right)^2} \\
&= \operatorname*{plim}_{T\to\infty} \frac{1}{T} \left(\sum \xi_t^2 + \beta^2 \sum z_t^2\right) = \sigma_\xi^2 + \beta^2 \sigma_z^2.
\end{aligned}$$

Finally, we analyze the asymptotic behavior of the t-statistics. The expressions for $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ are given in equation (3.30).

$$\operatorname*{plim}_{T\to\infty} t_a = \operatorname*{plim}_{T\to\infty} \hat{a} \left(\operatorname*{plim}_{T\to\infty} s^2\right)^{-1/2} \left(\operatorname*{plim}_{T\to\infty} \left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}_{(11)}\right)^{-1/2}$$

$$
\begin{aligned}
&= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\frac{1}{T}\hat{b}^{(d)}+\frac{1}{T}\chi^2 c}{\hat{b}^{(d)}}\right)^{-1/2}\\
&= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}{Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^4)}\right)^{-1/2}\\
T^{-1/2}\operatorname*{plim}_{T\to\infty}t_a &= \alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^{-3}c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(1)}{T^{-3}c^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(1)}\right)^{-1/2}=\alpha\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}.
\end{aligned}
$$

In a similar manner, we determine the asymptotics of t_b. To that end, we obtain an expression for $\mathbf{B}^{-1}$. We can write $\mathbf{B}=\sum\varepsilon_{t-1}^2\mathbf{p}\mathbf{p}'+\operatorname{diag}(\mathbf{v})+O_p(T^{1/2})$. Thus, we know that

$$
\mathbf{B}^{-1}=(1/f)\left(\left(\operatorname{diag}(\mathbf{v})\right)^{-1}f-\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\right)+O_p(T^{-1/2}),
$$

where $f=1+\sum\varepsilon_{t-1}^2\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\mathbf{p}$. Then

$$
\begin{aligned}
\operatorname*{plim}_{T\to\infty}t_b &= \operatorname*{plim}_{T\to\infty}\hat{b}\left(\operatorname*{plim}_{T\to\infty}s^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}_{(22)}\right)^{-1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{Tc^2}{\hat{b}^{(d)}}\right)^{-1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^3+O_p(T^2)}{Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^4)}\right)^{-1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{T^{-1}}{T^{-1}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2}\right)^{-1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2\right)^{1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\left(\sum\varepsilon_{t-1}^2\right)^2\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\mathbf{p}}{1+\sum\varepsilon_{t-1}^2\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\mathbf{p}}\right)^{1/2}\\
&= \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\operatorname*{plim}_{T\to\infty}\frac{\left(\sum\varepsilon_{t-1}^2\right)^2\sum_{k=1}^K\rho_k^2\left(\sum v_{k,t-1}^2\right)^{-1}}{1+\sum\varepsilon_{t-1}^2\sum_{k=1}^K\rho_k^2\left(\sum v_{k,t-1}^2\right)^{-1}}\right)^{1/2}
\end{aligned}
$$

$$T^{-1/2}\operatorname*{plim}_{T\to\infty} t_b = \beta\left(\sigma_\xi^2+\beta^2\sigma_z^2\right)^{-1/2}\left(\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}{1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\right)^{1/2}$$

If $\beta=0$: Note that the asymptotic behaviour of $\hat{a}$, $\hat{b}^{(d)}$, $s^{2(d)}$, $\operatorname*{plim}_{T\to\infty}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}_{(11)}$, and $\operatorname*{plim}_{T\to\infty}\left(\mathbf{X}'\mathbf{Q}\left[\mathbf{Q}'\mathbf{Q}\right]^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}_{(22)}$ remains unaltered. However, the convergence of $\hat{b}^{(n)}$ and $s^{2(n)}$ changes. For the former, we find the following

$$\begin{aligned}
\hat{b}^{(n)} &= \underbrace{c\mathbf{r}'Tc\mathbf{B}^{-1}\mathbf{t}-T\mathbf{r}'yc\mathbf{B}^{-1}\mathbf{q}}_{O_p(T^{7/2})}+o_p(T^3)\\
&= \underbrace{Tc\mathbf{r}'\mathbf{B}^{-1}\left(T\mathbf{p}\sum\xi_t\varepsilon_{t-1}+T\mathbf{w}\right)}_{O_p(T^{7/2})}+o_p(T^3)\\
&= \underbrace{T^3\sum\varepsilon_{t-1}^2\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum\xi_t\varepsilon_{t-1}+\mathbf{w}\right)}_{O_p(T^{7/2})}+O_p(T^3).
\end{aligned}$$

Hence, in the limit we get

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty}\hat{b} &= \operatorname*{plim}_{T\to\infty}\frac{T^3\sum\varepsilon_{t-1}^2\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum\xi_t\varepsilon_{t-1}+\mathbf{w}\right)+O_p(T^3)}{T^3\mathbf{p}'\sum\varepsilon_{t-1}^2\mathbf{B}^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}+o_p(T^4)}\\
&= \operatorname*{plim}_{T\to\infty}\frac{T^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}'\mathbf{B}^{-1}\left(\mathbf{p}\sum\xi_t\varepsilon_{t-1}+\mathbf{w}\right)}{T^{-1}\mathbf{p}'\sum\varepsilon_{t-1}^2\mathbf{B}^{-1}\sum\varepsilon_{t-1}^2\mathbf{p}}\\
&= \operatorname*{plim}_{T\to\infty}\frac{\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\left(\mathbf{p}\sum\xi_t\varepsilon_{t-1}+\mathbf{w}\right)}{\mathbf{p}'\left(\operatorname{diag}(\mathbf{v})\right)^{-1}\mathbf{p}\sum\varepsilon_{t-1}^2}\\
&= \operatorname*{plim}_{T\to\infty}\frac{\sum_{k=1}^K\rho_k^2\left(\sum v_{k,t-1}^2\right)^{-1}\left(\sum\xi_t\varepsilon_{t-1}+\frac{\sum\xi_t v_{k,t-1}}{\rho_k}\right)}{\sum_{k=1}^K\rho_k^2\left(\sum v_{k,t-1}^2\right)^{-1}\sum\varepsilon_{t-1}^2}\\
T^{1/2}\operatorname*{plim}_{T\to\infty}\hat{b} &= \operatorname*{plim}_{T\to\infty}\frac{\sum_{k=1}^K\rho_k^2\left(\frac{1}{T}\sum v_{k,t-1}^2\right)^{-1}\left(\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}+\frac{\frac{1}{T^{1/2}}\sum\xi_t v_{k,t-1}}{\rho_k}\right)}{\sum_{k=1}^K\rho_k^2\left(\frac{1}{T}\sum v_{k,t-1}^2\right)^{-1}\frac{1}{T}\sum\varepsilon_{t-1}^2}\\
&= \frac{\operatorname*{plim}_{T\to\infty}\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}}{\sigma_\varepsilon^2}+\frac{\sum_{k=1}^K\frac{\rho_k}{\sigma_{v_k}^2}\operatorname*{plim}_{T\to\infty}\frac{1}{T^{1/2}}\sum\xi_t v_{k,t-1}}{\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}.
\end{aligned}\tag{3.31}$$

Note that (3.31) is the sum of two independent random variables that has zero mean and asymptotic variance given by

$$\operatorname{Var}(T^{1/2}\operatorname*{plim}_{T\to\infty}\hat{b}) = \frac{\sigma_\xi^2\sigma_\varepsilon^2}{\sigma_\varepsilon^4} + \frac{\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{v_k}^4}\sigma_\xi^2\sigma_{v_k}^2}{\sigma_\varepsilon^4\left(\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{v_k}^2}\right)^2} = \frac{\sigma_\xi^2\left(\sigma_\varepsilon^2\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{v_k}^2}+1\right)}{\sigma_\varepsilon^4\sum_{k=1}^{K}\frac{\rho_k^2}{\sigma_{v_k}^2}}. \tag{3.32}$$

Thus, the result in Theorem 2 follows suit. Now we consider the error variance s^2. For the case where $\beta = 0$, we can write the numerator of s^2, $s^{2(n)}$, in the limit as

$$\begin{aligned}
s^{2(n)} &= \underbrace{\hbar\hat{b}^{(d)} - 2y\hat{a}^{(n)} + T\hat{a}\hat{a}^{(n)}}_{O_p(T^5)} + o_p(T^{9/2}) \\
\operatorname*{plim}_{T\to\infty} s^{2(n)} &= \operatorname*{plim}_{T\to\infty}\left\{\left(\alpha^2 T + \sum\xi_t^2\right)Tc^2\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} - 2\alpha Tc^2 y\mathbf{r}'\mathbf{B}^{-1}\mathbf{r} + T\alpha c^2 y\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\right\} \\
&= \operatorname*{plim}_{T\to\infty}\left\{\left(\alpha^2 T + \sum\xi_t^2\right)T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2 - 2\alpha^2 T^4\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2\right. \\
&\quad \left. + T^4\alpha^2\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2\right\} \\
&= \operatorname*{plim}_{T\to\infty}\left\{\sum\xi_t^2 T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2\right\}
\end{aligned}$$

Since the denominator is the same as above, we can conclude that

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} s^2 &= \operatorname*{plim}_{T\to\infty}\frac{\sum\xi_t^2 T^3\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2 + o_p(T^{9/2})}{T^4\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2 + o_p(T^5)} \\
&= \operatorname*{plim}_{T\to\infty}\frac{T^{-2}\sum\xi_t^2\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2}{T^{-1}\mathbf{p}'\mathbf{B}^{-1}\mathbf{p}\left(\sum\varepsilon_{t-1}^2\right)^2} = \sigma_\xi^2
\end{aligned} \tag{3.33}$$

In the final step, we analyze the asymptotic behavior of the t-statistics. Note that the computation of the limiting behavior of t_a follows exactly the steps in (3.31) in Section A above, with $\operatorname*{plim}_{T\to\infty} s^2$ replaced by σ_ξ^2. Then

$$T^{-1/2}\operatorname*{plim}_{T\to\infty} t_a = \alpha\left(\sigma_\xi^2\right)^{-1/2}.$$

For the derivation of t_b, we make the following replacements in equation (3.31) above.

Replace $\operatorname*{plim}_{T\to\infty} s^2$ by σ_ξ^2 and $\operatorname*{plim}_{T\to\infty} \hat{b}$ by $T^{-1/2}$ times the expression in (3.31). It follows that

$$T^{-1/2}\operatorname*{plim}_{T\to\infty} t_b = T^{-1/2}\frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2}\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}+\frac{\frac{1}{T^{1/2}}\sum\xi_t v_{k,t-1}}{\rho_k}\right)}{\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\left(\sigma_\xi^2\right)^{-1/2}\left(\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}{1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\right)^{1/2}$$

$$\operatorname*{plim}_{T\to\infty} t_b = \frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2}\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}+\frac{\frac{1}{T^{1/2}}\sum\xi_t v_{k,t-1}}{\rho_k}\right)}{\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\left(\sigma_\xi^2\right)^{-1/2}\left(\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}{1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\right)^{1/2} \tag{3.34}$$

By the same logic as before, note that (3.34) is a random variable with zero mean and asymptotic variance equal to

$$\operatorname{Var}\left\{\frac{\sum_{k=1}^K \frac{\rho_k^2}{\sigma_{v_k}^2}\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T^{1/2}}\sum\xi_t\varepsilon_{t-1}+\frac{\frac{1}{T^{1/2}}\sum\xi_t v_{k,t-1}}{\rho_k}\right)}{\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\right\}\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}\right)}$$

$$\frac{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}\right)}{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}\frac{\sigma_\varepsilon^4\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}}{\sigma_\xi^2\left(1+\sigma_\varepsilon^2\sum_{k=1}^K\frac{\rho_k^2}{\sigma_{v_k}^2}\right)} = 1,$$

where the last line follows directly from (3.32).

Proof of Corollary 1

This section presents proofs for the asymptotic results in Corollary 1. Sargan's $\mathscr{J}$ test for instrument validity is built upon a two-step procedure: (i) we estimate regression (3.6) by IV and, (ii) the resulting residuals, $\hat{\mathbf{e}}$, are in turn regressed on the instruments.

The residuals of the second regression, $\hat{\mathbf{v}}$, as well as $\hat{\mathbf{e}}$ are then used to construct the $\mathscr{J}$ statistic, $\mathscr{J} = T\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}-\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}}$, where $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathrm{b}}_{IV}$ and $\hat{\mathbf{v}} = \hat{\mathbf{e}} - \mathbf{Q}\hat{\varpi}$. Note that the test statistic can be written as

$$
\begin{aligned}
\mathscr{J} = T\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}-\hat{\mathbf{v}}'\hat{\mathbf{v}}}{\hat{\mathbf{e}}'\hat{\mathbf{e}}} &= \frac{\hat{\mathbf{e}}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\hat{\mathbf{e}}}{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}} \\
&= \frac{\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)'\mathbf{QL}}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{QLL}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{QL}\right]\frac{\mathbf{L}'\mathbf{Q}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}},
\end{aligned}
\tag{3.35}
$$

where

$$
\begin{aligned}
\mathbf{z}' &\equiv \begin{pmatrix} z_1 & z_2 & z_3 & \dots & z_{T-1} \end{pmatrix} \\
\boldsymbol{\xi}' &\equiv \begin{pmatrix} \xi_2 & \xi_3 & \xi_4 & \dots & \xi_T \end{pmatrix}
\end{aligned}
$$

and $\mathbf{L}$ is a $(K+1)\times(K+1)$ matrix such that $\mathbf{LL}' = (\mathbf{Q}'\mathbf{Q})^{-1}$. We can write $\mathbf{L}$ as

$$
\mathbf{L} = \begin{pmatrix} \frac{1}{\sqrt{c}} & 0 \\ -\frac{1}{\sqrt{c}}\mathbf{B}^{-1}\mathbf{q} & \mathbf{B}^{-1/2} \end{pmatrix}.
$$

$\mathscr{J}$ in (3.35) is the product of the transpose of a $(K+1)\times 1$ vector multiplied by a $(K+1)\times(K+1)$ symmetric and idempotent matrix, $\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{QLL}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{QL}\right]$, that has rank $K-1$, multiplied by the former $(K+1)\times 1$ vector. If it can be proven that the $(K+1)\times 1$ vector, $\frac{\mathbf{L}'\mathbf{Q}'(-\beta\mathbf{z}+\boldsymbol{\xi})}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}$, converges to a standard normal distribution, the usual result follows and $\mathscr{J} \xrightarrow{D} \chi^2_{(K-1)}$.

We can express the vector $\frac{\mathbf{L}'\mathbf{Q}'(-\beta\mathbf{z}+\boldsymbol{\xi})}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}$ as

$$
\frac{\mathbf{L}'\mathbf{Q}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}} = \frac{1}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\begin{pmatrix} -\frac{\beta}{\sqrt{c}}\sum z_{t-1} + \frac{1}{\sqrt{c}}\sum \xi_t + \frac{\beta}{\sqrt{c}}\mathbf{q}'\mathbf{B}^{-1}\mathbf{s} - \frac{1}{\sqrt{c}}\mathbf{q}'\mathbf{B}^{-1}\mathbf{a} \\ -\beta\mathbf{B}^{-1/2}\mathbf{s} + \mathbf{B}^{-1/2}\mathbf{a} \end{pmatrix},
\tag{3.36}
$$

where

$$
\mathbf{s} \equiv \begin{pmatrix} \sum z_{t-1} q_{1,t-1} \\ \vdots \\ \sum z_{t-1} q_{K,t-1} \end{pmatrix} \quad \mathbf{a} \equiv \begin{pmatrix} \sum \xi_t q_{1,t-1} \\ \vdots \\ \sum \xi_t q_{K,t-1} \end{pmatrix}.
$$

Now notice that the idempotent matrix that scales the vector in (3.36) can be re-written as

$$\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\right]=$$

$$\begin{pmatrix} \frac{\mathbf{q}'\mathbf{B}^{-1}\mathbf{q}}{T}-\frac{\mathbf{q}'\mathbf{B}^{-1}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1}\mathbf{q}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2c} & -\frac{\sqrt{c}}{T}\mathbf{q}'\mathbf{B}^{-1/2}+\frac{\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1/2}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2c} \\ -\frac{\sqrt{c}}{T}\mathbf{B}^{-1/2}\mathbf{q}+\frac{\sqrt{c}\mathbf{B}^{-1/2}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1}\mathbf{q}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2c} & \frac{\mathbf{B}^{-1/2}(T\mathbf{B}-\mathbf{q}\mathbf{q}')\mathbf{B}^{-1/2}}{T}-\frac{c\mathbf{B}^{-1/2}\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)\left(\frac{\chi}{\sqrt{T}}\mathbf{q}-\sqrt{T}\mathbf{r}\right)'\mathbf{B}^{-1/2}}{T(\chi-\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2+Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-\chi^2c} \end{pmatrix}.$$

It follows that the in the limit we find

$$\operatorname*{plim}_{T\to\infty}\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\right]=$$

$$\begin{pmatrix} O_p(T^{-1})\left\{\frac{\operatorname*{plim}_{T\to\infty}\left(c\mathbf{q}'\mathbf{B}^{-1}\mathbf{q}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-c(\mathbf{r}'\mathbf{B}^{-1}\mathbf{q})^2\right)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)} \right. & \left.\frac{\operatorname*{plim}_{T\to\infty}\left(T\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1/2}-T\sqrt{c}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1/2}\right)+o_p(T^2)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}\right\}O_p(T^{-\frac{1}{2}}) \\ O_p(T^{-\frac{1}{2}})\left\{\frac{\operatorname*{plim}_{T\to\infty}\left(T\sqrt{c}\mathbf{B}^{-1/2}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}-T\sqrt{c}\mathbf{B}^{-1/2}\mathbf{q}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\right)+o_p(T^2)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}\right. & \left.\frac{\operatorname*{plim}_{T\to\infty}\mathbf{B}^{-1/2}\left(T\mathbf{B}c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-Tc\mathbf{r}\mathbf{r}'\right)\mathbf{B}^{-1/2}+o_p(T^3)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}+o_p(T^3)}\right\}O_p(1) \end{pmatrix}$$

This implies that the limit of the idempotent matrix and the $(K+1)\times 1$ vector is equal to

$$\begin{aligned}
&\operatorname*{plim}_{T\to\infty}\left\{\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\right]\frac{\mathbf{L}'\mathbf{Q}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T}}}\right\}\\
&=\operatorname*{plim}_{T\to\infty}\left[\mathbf{I}-\mathbf{L}'\mathbf{Q}'\mathbf{X}\left(\mathbf{X}'\mathbf{Q}\mathbf{L}\mathbf{L}'\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{L}\right]\frac{\operatorname*{plim}_{T\to\infty}\mathbf{L}'\mathbf{Q}'\left(-\beta\mathbf{z}+\boldsymbol{\xi}\right)}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\\
&=\frac{1}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\begin{pmatrix}\frac{\operatorname*{plim}_{T\to\infty}\left(T\sqrt{c}\mathbf{q}'\mathbf{B}^{-1}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1/2}-T\sqrt{c}\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}\mathbf{q}'\mathbf{B}^{-1/2}\right)}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\operatorname*{plim}_{T\to\infty}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)\\
\frac{\operatorname*{plim}_{T\to\infty}\mathbf{B}^{-1/2}\left(T\mathbf{B}c\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}-Tc\mathbf{r}\mathbf{r}'\right)\mathbf{B}^{-1/2}}{\operatorname*{plim}_{T\to\infty}Tc\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\operatorname*{plim}_{T\to\infty}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)+o_p(1)\end{pmatrix}\\
&=\frac{1}{\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}}\begin{pmatrix}0\\
\operatorname*{plim}_{T\to\infty}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\operatorname*{plim}_{T\to\infty}\left(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a}\right)+o_p(1)\end{pmatrix};
\end{aligned}$$

i.e., the first row of the vector $[\mathbf{L}'\mathbf{Q}'(-\beta\mathbf{z}+\boldsymbol{\xi})]/\sqrt{\hat{\mathbf{e}}'\hat{\mathbf{e}}/T}$ in (3.36) disappears in the limit. The remaining terms, $[\operatorname*{plim}_{T\to\infty}(-\beta\mathbf{B}^{-1/2}\mathbf{s}+\mathbf{B}^{-1/2}\mathbf{a})]/\sqrt{\beta^2\sigma_z^2+\sigma_\xi^2}$, are (scaled) sums that by Lemma 1 converge to a standard normal distribution. Hence, $\mathscr{J}$ is asymptotically equivalent to

$$\begin{aligned}
&\begin{pmatrix}0 & \mathbf{n}'\end{pmatrix}\times\begin{pmatrix}0 & 0\\ 0 & \operatorname*{plim}_{T\to\infty}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\end{pmatrix}\times\begin{pmatrix}0\\ \mathbf{n}\end{pmatrix}\\
=\;&\mathbf{n}'\operatorname*{plim}_{T\to\infty}\left(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}\right)\mathbf{n}\sim\chi^2_{(K-1)},
\end{aligned}$$

where $\mathbf{n}$ is a $K\times 1$ vector that has a standard normal distribution, $\mathcal{N}(0,1)$. Since $\operatorname*{plim}_{T\to\infty}(I-\frac{\mathbf{B}^{-1/2}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{\mathbf{r}'\mathbf{B}^{-1}\mathbf{r}})$ is the probability limit of a $K\times K$ symmetric and idempotent matrix of rank $K-1$ the above result holds and it follows that

$$\mathscr{J}\xrightarrow{D}\chi^2_{(K-1)}.$$

B Figures

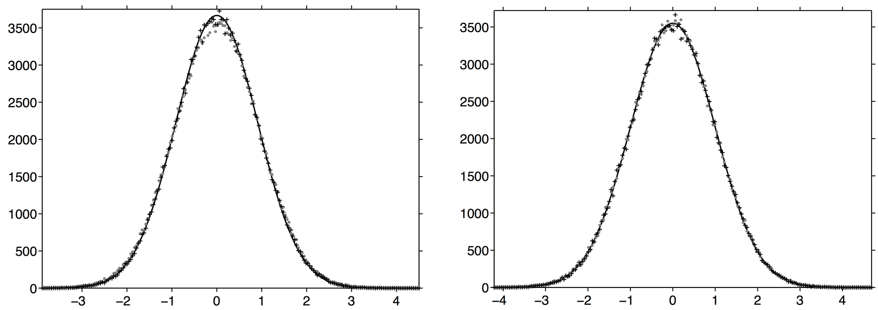


Figure 3.1. Small sample behavior of OLS estimates if $\beta = 0$ - The figures plot the small sample distribution of the scaled *OLS* estimate $T^{1/2}\hat{b}$ (left) from 200,000 simulations of the *DGP* (3.2)-(3.5), and the associated t-statistic, t_b (right). The black solid line reports the asymptotic distribution from Theorem 1. The gray dots represent the empirical distribution for $T = 250$; the black crosses are the empirical distribution for $T = 50,000$. In the simulations, we let $d = 0.35$, $\sigma_\eta \approx 1$, $\sigma_\xi \approx 2$, $\sigma_\varepsilon \approx 1.8$, $\alpha = 1.2$, and $\beta = 0$. The innovations in the *DGP* are drawn from continuous uniform distributions.

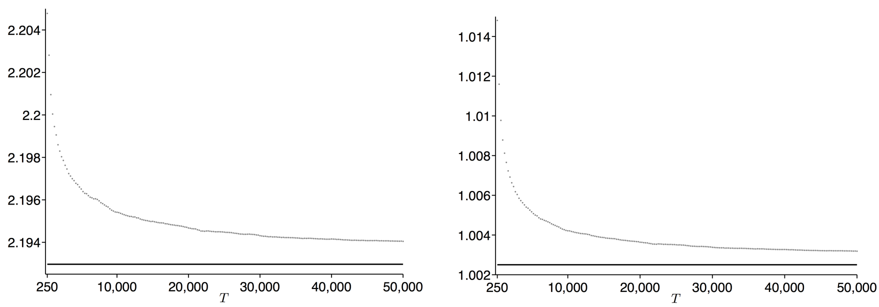


Figure 3.2. Small sample behavior of OLS estimates if $\beta \neq 0$ - The figures plot the small sample behavior of the *OLS* estimate $\hat{b}$ (left) from 200,000 simulations of the *DGP* (3.2)-(3.5), and the associated scaled t-statistic, $T^{-1/2}t_b$ (right). The x-axis contains varying sample sizes from $T = 250$ to $T = 50,000$. The black solid line reports the asymptotic value from Theorem 1. The gray dots represent the average estimate for a given T. In the simulations, we let $d = 0.2$, $\sigma_\eta \approx 1.2$, $\sigma_\xi \approx 1.7$, $\sigma_\varepsilon \approx 1.4$, $\alpha = 1.2$, and $\beta = 4$. The innovations in the *DGP* are drawn from t-distributions.

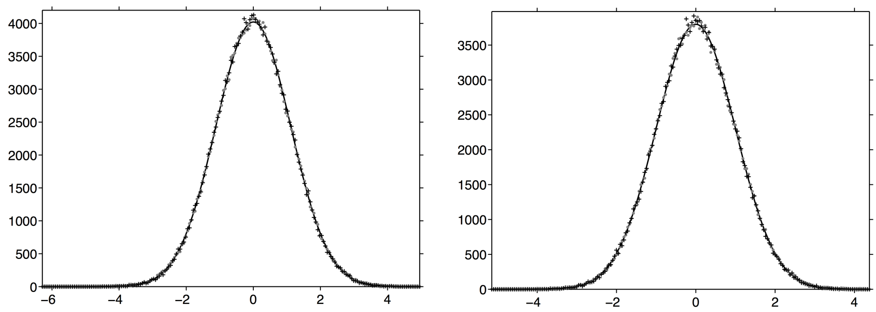


Figure 3.3. Small sample behavior of IV estimates if $\beta = 0$ - The figures plot the small sample distribution of the scaled IV estimate $T^{1/2}\hat{b}$ (left) from 200,000 simulations of the DGP (3.2)-(3.5) and the instruments (3.12) with $K = 9$, and the associated t-statistic, t_b (right). The black solid line reports the asymptotic distribution from Theorem 2. The gray dots represent the empirical distribution for $T = 250$; the black crosses are the empirical distribution for $T = 50{,}000$. In the simulations, we let $d = 0.3$, $\sigma_\eta = 1$, $\sigma_\xi = 2$, $\sigma_\varepsilon = 1.8$, $\sigma_\upsilon = [1.5, 1.2, 3.0, 1.5, 1.2, 3.0, 1.5, 1.2, 3.0]'$, $\alpha = 1.2$, $\beta = 0$, and $\rho = [3.77, 4.44, 1.77, 0.55, 3.11, 2.66, 1.99, 3.99, 0.99]'$. The innovations in the DGP are drawn from standard normal distributions.

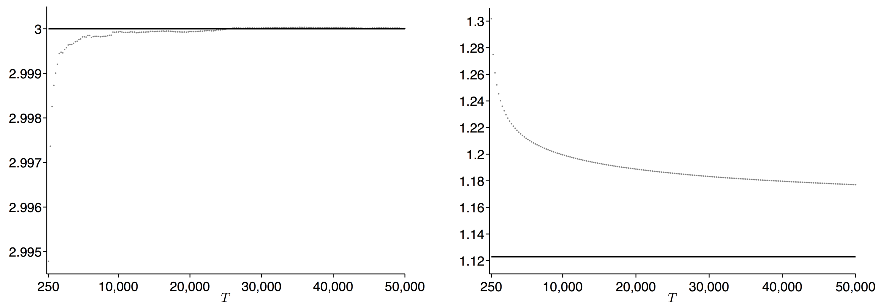


Figure 3.4. Small sample behavior of IV estimates if $\beta \neq 0$ - The figures plot the small sample behavior of the IV estimate $\hat{b}$ (left) from 200,000 simulations of the DGP (3.2)-(3.5) and the instruments (3.12) with $K = 3$, and the associated scaled t-statistic, $T^{-1/2}t_b$ (right). The x-axis contains varying sample sizes from $T = 250$ to $T = 50{,}000$. The black solid line reports the asymptotic value from Theorem 2. The gray dots represent the average estimate for a given T. In the simulations, we let $d = 0.4$, $\sigma_\eta \approx 1.0$, $\sigma_\xi \approx 2.0$, $\sigma_\varepsilon \approx 1.8$, $\sigma_\upsilon \approx [1.5, 1.2, 3.0]'$, $\alpha = 1.2$, $\beta = 3$, and $\rho = [3.77, 4.44, 1.77]'$. The innovations in the DGP are drawn from continuous uniform distributions.

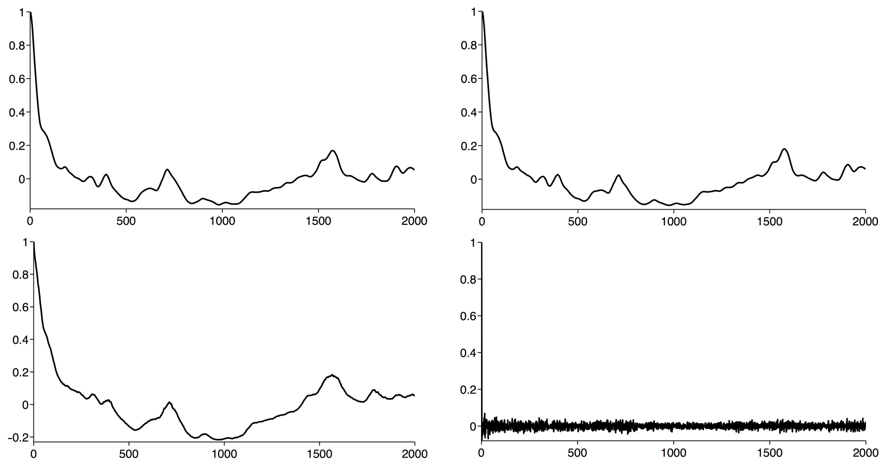


Figure 3.5. ACF estimates for the three variance series and returns - The figure plots the estimates of the autocorrelation of the realized variance, rv_t (top left), the bipower variation, bv_t (top right), the volatility index, vix_t^2 (bottom left), and daily intraday returns on the the S&P 500, r_t (bottom right). The x-axis measures lags in daily units.

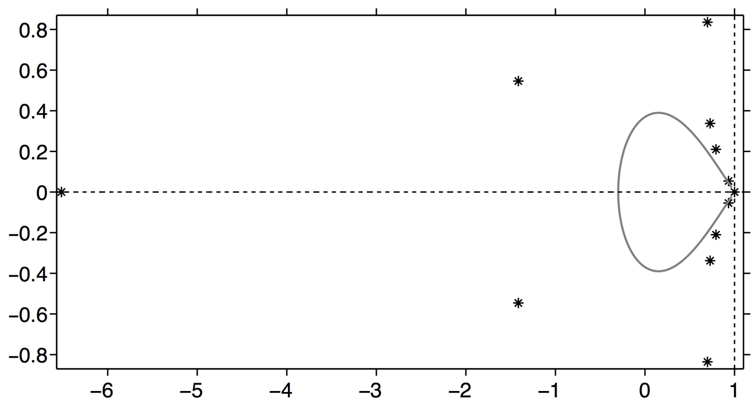


Figure 3.6. Roots of the characteristic polynomial of the co-fractional *VAR* - The figure plots the roots of the characteristic equation $|(1-c)I_{3\times3} - \varphi\theta' c - (1-c)\sum_{i=1}^{n}\Gamma_i c^i| = 0$, indicated by the black stars. The gray line is the image of the complex disk $\mathbb{C}_d$, for $\hat{d} = 0.3775$. For $\theta' X_t$ to be $I(0)$, all roots must be equal to one or lie outside the disk.

C Tables

Table 3.3. Size, Power, and (In-)Consistency of *OLS* Estimate $\hat{b}$

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0: \beta = 0$ vs $H_1: \beta \neq 0$. We estimate regression (3.6) by *OLS*. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from t-distributions.

		d															
		β															
		0.1				0.23				0.36				0.49			
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta$	T	-3	0	0.8	3	-3	0	0.8	3	-3	0	0.8	3	-3	0	0.8	3
1.7,1.7,1.7	250	99	5.1	94	99	99	5.1	92	99	99	5.1	83	98	90	5.1	48	90
		50		50	50	48		48	48	44		44	44	36		36	36
1.7,1.7,1.7	1000	100	5.0	100	100	100	5.0	100	100	100	5.1	99	100	99	5.0	94	99
		50		50	50	47		47	47	42		42	42	32		32	32
1.7,1.7,1.1	250	100	5.1	100	100	100	5.2	100	100	100	5.2	99	100	100	5.3	93	100
		68		68	68	66		66	66	62		62	62	54		54	54
1.7,1.7,1.1	1000	100	5.0	100	100	100	5.0	100	100	100	5.0	100	100	100	5.0	100	100
		69		69	69	66		66	66	61		61	61	50		50	50
1.7,1.1,1.7	250	98	5.1	64	98	98	5.1	57	98	95	5.1	35	95	67	5.2	9	67
		32		32	32	30		30	30	27		27	27	21		21	21
1.7,1.1,1.7	1000	100	5.1	99	100	100	5.1	98	100	100	5.0	95	100	97	5.0	48	97
		31		31	31	29		29	29	24		24	24	17		17	17
1.7,1.1,1.1	250	100	5.1	97	100	100	5.1	96	100	100	5.1	91	100	99	5.1	59	99
		50		50	50	48		48	48	43		43	43	36		36	36
1.7,1.1,1.1	1000	100	5.0	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	98	100
		50		50	50	47		47	47	41		41	41	31		31	31
1.1,1.7,1.7	250	99	5.2	98	99	99	5.1	98	99	99	5.1	95	99	92	5.1	72	92
		50		50	50	48		48	48	44		44	44	36		36	36
1.1,1.7,1.7	1000	100	4.9	100	100	100	5.0	100	100	100	5.0	100	100	99	5.1	98	99
		50		50	50	47		47	47	42		42	42	32		32	32
1.1,1.7,1.1	250	100	5.1	100	100	100	5.2	100	100	100	5.2	100	100	100	5.2	99	100
		68		68	68	66		66	66	62		62	62	54		54	54
1.1,1.7,1.1	1000	100	5.0	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	100	100
		69		69	69	66		66	66	61		61	61	50		50	50
1.1,1.1,1.7	250	99	5.1	90	99	98	5.2	86	98	96	5.1	69	96	73	5.2	26	73
		32		32	32	30		30	30	27		27	27	21		21	21
1.1,1.1,1.7	1000	100	5.0	100	100	100	5.1	99	100	100	5.1	99	100	98	5.1	80	98
		31		31	31	29		29	29	24		24	24	17		17	17
1.1,1.1,1.1	250	100	5.0	100	100	100	5.1	100	100	100	5.2	100	100	100	5.1	88	100
		50		50	50	48		48	48	43		43	43	36		36	36
1.1,1.1,1.1	1000	100	5.1	100	100	100	5.1	100	100	100	5.0	100	100	100	5.1	100	100
		50		50	50	47		47	47	41		41	41	31		31	31

Table 3.4. Size, Power, and Consistency of IV Estimate $\hat{b}$

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate regression (3.6) by IV. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from standard normal distributions.

		d																	
		Corr(q_t, x_t^*)																	
		β																	
		0.1						0.295						0.49					
		-0.8			0.65			-0.8			0.65			-0.8			0.65		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\upsilon$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	100	4.7	100	100	4.4	100	100	4.7	100	100	4.3	100	100	4.5	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.7,1.7,1.7	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.7,1.7,1.1	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.3	100	100	4.5	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.7,1.7,1.1	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.7,1.1,1.7	250	100	4.9	100	100	4.6	100	100	4.8	100	100	4.7	100	100	4.8	100	100	4.5	100
		100		100	100		100	100		100	100		100	100		101	101		101
1.7,1.7,1.1,1.7	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	5.0	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.7,1.7,1.1,1.1	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.6	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.7,1.7,1.1,1.1	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.7,1.1,1.7,1.7	250	100	4.3	100	100	3.8	100	100	4.2	100	100	3.6	100	100	3.8	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.7,1.1,1.7,1.7	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.7	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.7,1.1,1.7,1.1	250	100	4.3	100	100	3.8	100	100	4.1	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.7,1.1,1.7,1.1	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.7,1.1,1.1,1.7	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.1,1.1,1.7	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.7,1.1,1.1,1.1	250	100	4.6	100	100	4.5	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.7,1.1,1.1,1.1	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.7,1.7	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.3	100	100	4.5	100	100	4.0	100

Table 3.4 – continued from previous page

		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.7,1.7,1.7	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.9	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.7,1.1	250	100	4.7	100	100	4.4	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.7,1.7,1.1	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.7,1.1,1.7	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.6	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.1,1.7,1.1,1.7	1000	100	5.0	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100	100	4.8	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.1,1.7,1.1,1.1	250	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.7	100	100	4.5	100
		100		100	100		100	100		100	100		100	101		101	101		101
1.1,1.7,1.1,1.1	1000	100	5.0	100	100	4.9	100	100	4.9	100	100	5.0	100	100	4.9	100	100	4.9	100
		100		100	100		100	100		100	100		100	100		100	100		100
1.1,1.1,1.7,1.7	250	100	4.3	100	100	3.8	100	100	4.1	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.1,1.1,1.7,1.7	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.6	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.1,1.1,1.7,1.1	250	100	4.2	100	100	3.8	100	100	4.2	100	100	3.7	100	100	3.7	100	99	3.0	99
		102		102	102		102	102		102	103		103	103		103	105		105
1.1,1.1,1.7,1.1	1000	100	4.8	100	100	4.7	100	100	4.8	100	100	4.7	100	100	4.6	100	100	4.4	100
		100		100	101		101	100		100	101		101	101		101	101		101
1.1,1.1,1.1,1.7	250	100	4.6	100	100	4.5	100	100	4.6	100	100	4.4	100	100	4.4	100	100	4.0	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.1,1.1,1.7	1000	100	4.9	100	100	4.8	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101
1.1,1.1,1.1,1.1	250	100	4.7	100	100	4.5	100	100	4.6	100	100	4.3	100	100	4.4	100	100	4.1	100
		101		101	101		101	101		101	101		101	101		101	102		102
1.1,1.1,1.1,1.1	1000	100	4.9	100	100	4.9	100	100	4.9	100	100	4.8	100	100	4.8	100	100	4.7	100
		100		100	100		100	100		100	100		100	100		100	101		101

Table 3.5. Size, Power, and (In-)Consistency of IV Estimate $\hat{b}$ with Irrelevant Instrument

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate regression (3.6) by IV using an irrelevant instrument. That is, we set $\text{Corr}(q_t, x_t^*)$ equal to zero. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from continuous uniform distributions.

					d					
					β					
			0.1			0.295			0.49	
$\sigma_\xi,\sigma_\varepsilon,\sigma_\eta,\sigma_v$	T	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	4	0.01	5	3	0.01	4	2	0.01	3
		-218		-135	73		25	86		80
1.7,1.7,1.7,1.7	1000	4	0.01	5	3	0.01	4	2	0.01	2
		23		127	33		58	-84		6
1.7,1.7,1.7,1.1	250	4	0.01	5	3	0.02	5	2	0.01	3
		-122		-130	71		-1522	157		133
1.7,1.7,1.7,1.1	1000	4	0.01	5	3	0.01	4	2	0.01	2
		-526		278	25		46	-105		-80
1.7,1.7,1.1,1.7	250	9	0.01	13	8	0.01	11	6	0.01	7
		-412		1472	53		71	434		81
1.7,1.7,1.1,1.7	1000	9	0.01	13	8	0.01	11	4	0.01	6
		76		-414	-10		-228	412		-91
1.7,1.7,1.1,1.1	250	9	0.01	13	8	0.01	11	6	0.01	7
		-539		429	74		-27	121		406
1.7,1.7,1.1,1.1	1000	9	0.01	12	8	0.01	11	4	0.01	6
		-335		86	-56		294	483		54
1.7,1.1,1.7,1.7	250	1	0.01	1	1	0.01	1	0	0.01	1
		-13		18	358		-101	162		76
1.7,1.1,1.7,1.7	1000	1	0.01	1	1	0.01	1	0	0.01	0
		-294		65	-2227		-2268	147		7
1.7,1.1,1.7,1.1	250	1	0.01	1	1	0.01	1	0	0.01	1
		-2		-13	-73		-77	81		-10
1.7,1.1,1.7,1.1	1000	1	0.01	1	1	0.01	1	0	0.01	0
		-333		108	-2316		-2307	57		11
1.7,1.1,1.1,1.7	250	2	0.01	4	2	0.01	3	1	0.01	2
		-70		95	-66		-1928	247		71
1.7,1.1,1.1,1.7	1000	2	0.01	4	2	0.01	3	1	0.01	2
		148		59	68		23	-147		103
1.7,1.1,1.1,1.1	250	2	0.01	4	2	0.01	3	1	0.01	2
		140		426	-134		1974	27		-5
1.7,1.1,1.1,1.1	1000	2	0.01	4	2	0.01	3	1	0.01	1
		17		56	-10		-142	28		91
1.1,1.7,1.7,1.7	250	5	0.01	6	4	0.02	5	3	0.01	3
		-125		-124	61		-1367	141		145
1.1,1.7,1.7,1.7	1000	5	0.01	6	4	0.01	5	2	0.01	2
		-391		265	32		48	-82		-87
1.1,1.7,1.7,1.1	250	5	0.01	6	5	0.01	5	3	0.01	3
		-94		106	-633		-1419	201		61
1.1,1.7,1.7,1.1	1000	5	0.01	6	4	0.01	5	2	0.01	2
		200		49	59		16	-119		86
1.1,1.7,1.1,1.7	250	13	0.01	15	11	0.01	13	7	0.01	9
		-122		251	74		-8	113		408
1.1,1.7,1.1,1.7	1000	13	0.01	15	11	0.01	12	6	0.01	7
		-352		95	-92		262	364		48
1.1,1.7,1.1,1.1	250	13	0.01	15	11	0.01	13	7	0.01	9
		-593		-179	80		81	415		16
1.1,1.7,1.1,1.1	1000	13	0.01	15	11	0.01	12	6	0.01	7
		139		146	117		446	17		-278
1.1,1.1,1.7,1.7	250	1	0.01	2	1	0.01	1	1	0.01	1
		2		-14	-79		-70	80		-14
1.1,1.1,1.7,1.7	1000	1	0.01	2	1	0.01	1	0	0.01	1
		-250		66	-2306		-2332	47		18
1.1,1.1,1.7,1.1	250	1	0.01	2	1	0.01	1	1	0.01	1
		-17		-5	-37		-470	-33		-36

Table 3.5 – continued from previous page

1.1,1.1,1.7,1.1	1000	1	0.01	2	1	0.01	1	0	0.01	1
		-133		99	-2452		-9	47		120
1.1,1.1,1.1,1.7	250	4	0.01	5	3	0.01	4	2	0.01	3
		131		351	-506		1839	36		-11
1.1,1.1,1.1,1.7	1000	4	0.01	5	3	0.01	4	2	0.01	2
		26		51	-3		-134	43		85
1.1,1.1,1.1,1.1	250	4	0.01	5	3	0.01	4	2	0.01	3
		-3		426	1202		26	-38		27
1.1,1.1,1.1,1.1	1000	4	0.01	5	3	0.01	4	2	0.01	2
		27		-327	-93		53	58		-11

Table 3.6. Size, Power, and (In-)Consistency of IV Estimate $\hat{b}$ with Invalid Instruments

The table reports rejection rates in % at a nominal size of 5% based on a standard t-test of $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. We estimate regression (3.6) by IV using an invalid instrument of type 1 and type 2. That is, we set $\text{Corr}(q_{t-1}, z_{t-1}) = 0.7$ for the invalid instrument of type 1; for the invalid instrument of type 2, we let $\text{Corr}(q_{t-1}, \xi_t) = 0.7$. Both instruments are relevant with $\text{Corr}(q_t, x_t^*) = 0.7$. Simulations are based on 200,000 repetitions. The table also outlines the relative bias in the estimate, $(\hat{b}/\beta) \times 100$ in gray font. All errors are drawn from standard normal distributions.

		d Invalid Instrument β																	
		0.1						0.295						0.49					
		type 1			type 2			type 1			type 2			type 1			type 2		
$\sigma_\xi,\sigma_\varepsilon,\sigma_\eta,\sigma_\mu$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.7,1.7,1.7,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134
1.7,1.7,1.7,1.1	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.7,1.7,1.7,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134
1.7,1.7,1.1,1.7	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.1	99	100	100	93
		60		60	50		134	60		60	50		134	77		77	50		135
1.7,1.7,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	100
		60		60	50		134	59		59	50		134	74		74	50		134
1.7,1.7,1.1,1.1	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.1	99	100	100	94
		60		60	50		134	60		60	50		134	77		77	50		135
1.7,1.7,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	100
		60		60	50		134	59		59	50		134	74		74	50		134
1.7,1.1,1.7,1.7	250	100	5.1	100	100	100	80	100	5.1	100	100	100	68	100	4.9	100	99	100	22
		40		40	23		155	39		39	24		155	60		60	24		158
1.7,1.1,1.7,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		39		39	24		152	38		38	24		152	55		55	24		153
1.7,1.1,1.7,1.1	250	100	5.1	100	100	100	80	100	5.1	100	100	100	68	100	4.9	100	99	100	22
		40		40	23		155	39		39	24		155	60		60	24		158
1.7,1.1,1.7,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		39		39	24		152	38		38	24		152	55		55	24		153
1.7,1.1,1.1,1.7	250	100	5.1	100	100	100	100	100	5.1	100	100	100	99	99	5.0	100	100	100	90
		50		50	23		153	49		49	23		153	69		69	23		154
1.7,1.1,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.1	100	100	100	100	100	5.0	100	100	100	100
		50		50	23		151	48		48	24		151	65		65	24		152
1.7,1.1,1.1,1.1	250	100	5.1	100	100	100	100	100	5.2	100	100	100	99	99	5.0	100	100	100	90
		50		50	23		153	49		49	23		153	69		69	23		154
1.7,1.1,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	100

Table 3.6 – continued from previous page

		50		50	23		151	48		48	24		151	65		65	24		152
1.1,1.7,1.7,1.7	250	100	5.1	100	96	100	46	100	5.1	100	94	100	37	100	5.0	100	85	100	17
		50		50	68		123	49		49	68		123	69		69	68		124
1.1,1.7,1.7,1.7	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.0	100	100	100	83
		50		50	67		122	48		48	67		122	65		65	68		122
1.1,1.7,1.7,1.1	250	100	5.1	100	96	100	46	100	5.1	100	94	100	37	100	5.0	100	85	100	17
		50		50	68		123	49		49	68		123	69		69	68		124
1.1,1.7,1.7,1.1	1000	100	5.0	100	100	100	99	100	5.1	100	100	100	98	100	5.0	100	100	100	83
		50		50	67		122	48		48	67		122	65		65	68		122
1.1,1.7,1.1,1.7	250	100	5.1	100	100	100	90	100	5.1	100	100	100	84	99	5.1	100	98	100	62
		60		60	67		122	60		60	67		123	77		77	67		123
1.1,1.7,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.0	100	100	100	99
		60		60	67		122	59		59	67		122	74		74	67		122
1.1,1.7,1.1,1.1	250	100	5.1	100	100	100	90	100	5.1	100	100	100	84	100	5.1	100	98	100	62
		60		60	67		122	60		60	67		123	77		77	67		123
1.1,1.7,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	99
		60		60	67		122	59		59	67		122	74		74	67		122
1.1,1.1,1.7,1.7	250	100	5.1	100	97	100	30	100	5.1	100	96	100	18	100	4.9	100	89	100	2
		40		40	51		136	39		39	51		137	60		60	52		139
1.1,1.1,1.7,1.7	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.0	100	100	100	78
		39		39	50		134	38		38	50		134	55		55	50		135
1.1,1.1,1.7,1.1	250	100	5.1	100	97	100	30	100	5.1	100	96	100	18	100	4.9	100	89	100	2
		40		40	51		137	39		39	51		137	60		60	52		139
1.1,1.1,1.7,1.1	1000	100	5.0	100	100	100	99	100	5.0	100	100	100	98	100	5.1	100	100	100	78
		39		39	50		134	38		38	50		134	55		55	50		135
1.1,1.1,1.1,1.7	250	100	5.1	100	100	100	87	100	5.2	100	100	100	79	100	5.0	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.1,1.1,1.1,1.7	1000	100	5.0	100	100	100	100	100	5.0	100	100	100	100	100	5.1	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134
1.1,1.1,1.1,1.1	250	100	5.1	100	100	100	87	100	5.1	100	100	100	79	100	4.9	100	99	100	51
		50		50	50		135	49		49	50		135	69		69	50		136
1.1,1.1,1.1,1.1	1000	100	5.0	100	100	100	100	100	5.1	100	100	100	100	100	5.0	100	100	100	99
		50		50	50		134	48		48	50		134	65		65	50		134

Table 3.7. Size and Power of $\mathscr{J}$−test

The table reports rejection rates in % at a nominal size of 5% based on a $\mathscr{J}$-test of H_0 : valid instruments vs. H_1 : invalid instruments of type 2. We estimate regression (3.6) by IV with $K = 2$ and subsequently estimate regression (3.16) by *OLS* to compute the $\mathscr{J}$-statistic. $q_{1,t}$ and $q_{2,t}$ are strongly and weakly relevant, respectively, with $\text{Corr}([q_{1,t-1}\ q_{2,t-1}]', x^*_{t-1}) = [0.85, 0.1]'$. Simulations are based on 200,000 repetitions. All errors are drawn from t-distributions.

		d																	
		$\text{Corr}([q_{1,t-1}\ q_{2,t-1}]', \xi_t)$																	
		β																	
		0.1									0.49								
		$[0.5, -0.6]'$			$[0, 0]'$			$[-0.4, 0.3]'$			$[0.5, -0.6]'$			$[0, 0]'$			$[-0.4, 0.3]'$		
$\sigma_\xi, \sigma_\varepsilon, \sigma_\eta, \sigma_\mu$	T	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3	-2	0	3
1.7,1.7,1.7,1.7	250	99	100	76	5.3	5.1	5.3	52	97	55	94	100	57	5.4	5.1	5.4	37	95	40
1.7,1.7,1.7,1.7	1000	100	100	99	5.1	5.0	5.1	96	100	96	100	100	94	5.1	4.9	5.0	78	100	80
1.7,1.7,1.7,1.1	250	99	100	75	5.2	5.0	5.3	49	98	52	94	100	55	5.3	5.0	5.4	35	95	38
1.7,1.7,1.7,1.1	1000	100	100	100	5.1	5.0	5.0	96	100	96	100	100	94	5.2	4.9	5.2	77	100	79
1.7,1.7,1.1,1.7	250	100	100	94	5.1	5.1	5.2	74	98	75	99	100	82	5.2	5.1	5.2	57	97	60
1.7,1.7,1.1,1.7	1000	100	100	100	5.1	4.9	5.1	100	100	100	100	100	100	5.1	5.0	5.1	96	100	96
1.7,1.7,1.1,1.1	250	100	100	94	5.2	5.1	5.2	71	99	73	100	100	80	5.1	5.1	5.3	54	98	57
1.7,1.7,1.1,1.1	1000	100	100	100	5.1	5.0	5.1	100	100	100	100	100	100	5.0	5.1	5.1	96	100	96
1.7,1.1,1.7,1.7	250	99	100	73	5.5	5.0	5.5	48	95	59	95	99	54	5.7	4.9	5.9	35	90	45
1.7,1.1,1.7,1.7	1000	100	100	99	5.1	4.9	5.1	93	100	97	100	100	92	5.2	4.9	5.2	72	100	85
1.7,1.1,1.7,1.1	250	99	100	71	5.4	5.0	5.4	45	95	57	95	100	52	5.7	4.9	5.8	32	90	43
1.7,1.1,1.7,1.1	1000	100	100	99	5.1	5.0	5.1	93	100	97	100	100	92	5.2	5.0	5.2	71	100	84
1.7,1.1,1.1,1.7	250	100	100	93	5.1	5.0	5.3	70	98	79	100	100	78	5.3	5.0	5.3	53	96	64
1.7,1.1,1.1,1.7	1000	100	100	100	5.1	5.0	5.1	99	100	100	100	100	99	5.1	5.0	5.0	93	100	98
1.7,1.1,1.1,1.1	250	100	100	92	5.1	5.0	5.2	67	98	77	100	100	77	5.3	5.0	5.4	50	97	62
1.7,1.1,1.1,1.1	1000	100	100	100	5.1	5.0	5.0	100	100	100	100	100	100	5.1	5.0	5.0	94	100	98
1.1,1.7,1.7,1.7	250	94	100	56	5.2	5.1	5.2	37	98	32	81	100	37	5.5	5.0	5.5	25	95	22
1.1,1.7,1.7,1.7	1000	100	100	97	5.1	5.1	5.0	85	100	79	99	100	79	5.1	5.0	5.1	57	100	50
1.1,1.7,1.7,1.1	250	94	100	54	5.2	5.1	5.2	34	98	30	81	100	36	5.4	4.9	5.4	23	95	21
1.1,1.7,1.7,1.1	1000	100	100	97	5.1	5.1	5.1	84	100	78	99	100	78	5.1	5.0	5.2	56	100	49
1.1,1.7,1.1,1.7	250	100	100	83	5.2	5.1	5.2	59	99	54	97	100	62	5.1	5.0	5.2	42	98	37
1.1,1.7,1.1,1.7	1000	100	100	100	5.0	5.0	5.0	98	100	97	100	100	97	5.0	5.0	4.9	86	100	80
1.1,1.7,1.1,1.1	250	100	100	82	5.1	5.2	5.2	56	99	51	97	100	60	5.3	5.1	5.3	39	99	34
1.1,1.7,1.1,1.1	1000	100	100	100	5.0	5.0	5.0	99	100	97	100	100	98	5.1	5.0	5.1	85	100	79
1.1,1.1,1.7,1.7	250	95	100	54	5.5	4.9	5.4	34	96	35	83	99	36	5.7	4.9	5.8	24	91	24
1.1,1.1,1.7,1.7	1000	100	100	96	5.1	5.1	5.0	81	100	83	100	100	75	5.3	4.9	5.2	52	100	55
1.1,1.1,1.7,1.1	250	96	100	51	5.4	4.9	5.5	32	96	33	83	100	35	5.7	5.0	5.8	22	90	23
1.1,1.1,1.7,1.1	1000	100	100	96	5.0	5.0	5.0	80	100	82	100	100	74	5.2	4.9	5.2	51	100	53
1.1,1.1,1.1,1.7	250	100	100	81	5.3	5.1	5.3	56	98	57	98	100	60	5.4	5.1	5.4	39	97	40
1.1,1.1,1.1,1.7	1000	100	100	100	5.1	5.0	5.1	97	100	98	100	100	96	5.0	5.0	5.1	81	100	84

Table 3.8. Long-Memory Estimates

The upper panel of the table reports estimates of d using the multivariate EW estimator of Nielsen and Shimotsu (2007) for $Y_t = [rv_t,\ bv_t,\ vix_t^2,\ r_t]'$. The size of the spectral window is set to $m = T^{0.35}$; the choice is based on a graphical analysis of the slope of the log periodograms as suggested by Beran (1994). $t_{d=0}$ denotes the respective t-statistic of element i of Y_t given by $2\sqrt{m}\hat{d}_i$. The lower panel of the table summarizes the t-statistics corresponding to the null hypothesis $d_i = d_j$ for $i \neq j$. Nielsen and Shimotsu (2007) define the t-statistic as

$$t_{d_i=d_j} = \frac{\sqrt{m}\left(\hat{d}_i - \hat{d}_j\right)}{\sqrt{\frac{1}{2}\left(1 - \frac{\hat{\tau}_{i,j}^2}{\hat{\tau}_{i,i}\hat{\tau}_{j,j}}\right) + h(T)}},$$

where $\hat{\tau}_{i,j} = \frac{1}{m}\sum_{l=1}^{m} \text{real}\left\{I(\lambda_l)\right\}$ and $I(\lambda_l)$ is the periodogram of a (4×1) vector with elements $\Delta^{\hat{d}_i} Y_{t,i}$ at frequency λ_l. $h(T)$ is a tuning parameter, which we set equal to $\left(\ln(T)\right)^{-1}$ as in Nielsen and Shimotsu (2007). The resulting statistic $t_{d_i=d_j}$ should be compared to critical values from a t-distribution.

	Estimates for d			
	rv_t	bv_t	vix_t^2	r_t
$\hat{d}$	0.3517	0.3403	0.4393	0.0202
$t_{d=0}$	2.8134	2.7227	3.5146	0.1618

	$t_{d_i=d_j}$ statistics with $h(T) = 0.1233$			
	rv_t	bv_t	vix_t^2	r_t
rv_t	-	0.2609	-1.6730	2.2602
bv_t		-	-1.7115	2.1587
vix_t^2			-	2.9844
r_t				-

Table 3.9. Summary Statistics and Estimation Results

The first panel of the table reports summary statistics of the three variance series and intraday returns. The second panel summarizes the estimation results when the predictive regression (3.17) is evaluated by OLS. $OLS-SE$ denotes the usual standard error of b, and $HAC-SE$ reports standard errors based on HAC covariance estimation using a Bartlett kernel. The third panel of the table contains the analogous results from IV estimation. $IV-SE$ is the usual standard error of b and $\mathscr{J}$ is Sargan's statistic from Corollary 1.

Summary Statistics						
			Autocorrelation			
	Average	Std. Dev.	1	2	3	22
r_t	0.0139	1.2833	-0.0769	-0.0612	0.0205	0.0356
rv_t	31.7888	47.6128	0.9976	0.9927	0.9860	0.7009
bv_t	25.4352	40.5219	0.9976	0.9927	0.9858	0.6959
vix_t^2	45.9689	48.6179	0.9690	0.9469	0.9322	0.7413

OLS Regressions (3.17)				
x_t	$\hat{b}$	$OLS-SE(b)$	$HAC-SE(b)$	$t_{\hat{b}_{OLS}}-SE$
rv_t	4.98×10^{-5}	0.0005	0.0004	0.0996
bv_t	7.70×10^{-5}	0.0005	0.0004	0.1540
vix_t^2	0.0015	0.0005	0.0002	3.0012

IV Regressions (3.17)					
x_t	$\hat{b}$	$IV-SE(b)$	$HAC-SE(b)$	$t_{\hat{b}_{IV}}-SE$	$\mathscr{J}$
rv_t	-0.0130	0.0024	0.0046	-5.4166	13.7306
bv_t	-0.0088	0.0021	0.0035	-4.1905	30.3321
vix_t^2	0.0128	0.0020	0.0060	6.4001	1.4128

DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY
SCHOOL OF BUSINESS AND SOCIAL SCIENCES
www.econ.au.dk

PhD dissertations since 1 July 2011

2011-4	Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics
2011-5	Christian Bach: The Game of Risk
2011-6	Stefan Holst Bache: Quantile Regression: Three Econometric Studies
2011:12	Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management
2011:13	Christian Gormsen Schmidt: Exploring the Barriers to Globalization
2011:16	Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders
2011:22	Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade
2012-1	Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting
2012-2	Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping
2012-3	Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing
2012-4	Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing
2012-5	Lasse Frisgaard Gunnersen: Income Redistribution Policies
2012-6	Miriam Wüst: Essays on early investments in child health
2012-7	Yukai Yang: Modelling Nonlinear Vector Economic Time Series
2012-8	Lene Kjærsgaard: Empirical Essays of Active Labor Market Policy on Employment
2012-9	Henrik Nørholm: Structured Retail Products and Return Predictability
2012-10	Signe Frederiksen: Empirical Essays on Placements in Outside Home Care
2012-11	Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing

2012-12 Jens Riis Andersen: Option Games under Incomplete Information

2012-13 Margit Malmmose: The Role of Management Accounting in New Public Management Reforms: Implications in a Socio-Political Health Care Context

2012-14 Laurent Callot: Large Panels and High-dimensional VAR

2012-15 Christian Rix-Nielsen: Strategic Investment

2013-1 Kenneth Lykke Sørensen: Essays on Wage Determination

2013-2 Tue Rauff Lind Christensen: Network Design Problems with Piecewise Linear Cost Functions

2013-3 Dominyka Sakalauskaite: A Challenge for Experts: Auditors, Forensic Specialists and the Detection of Fraud

2013-4 Rune Bysted: Essays on Innovative Work Behavior

2013-5 Mikkel Nørlem Hermansen: Longer Human Lifespan and the Retirement Decision

2013-6 Jannie H.G. Kristoffersen: Empirical Essays on Economics of Education

2013-7 Mark Strøm Kristoffersen: Essays on Economic Policies over the Business Cycle

2013-8 Philipp Meinen: Essays on Firms in International Trade

2013-9 Cédric Gorinas: Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective

2013-10 Ina Charlotte Jäkel: Product Quality, Trade Policy, and Voter Preferences: Essays on International Trade

2013-11 Anna Gerstrøm: World Disruption - How Bankers Reconstruct the Financial Crisis: Essays on Interpretation

2013-12 Paola Andrea Barrientos Quiroga: Essays on Development Economics

2013-13 Peter Bodnar: Essays on Warehouse Operations

2013-14 Rune Vammen Lesner: Essays on Determinants of Inequality

2013-15 Peter Arendorf Bache: Firms and International Trade

2013-16 Anders Laugesen: On Complementarities, Heterogeneous Firms, and International Trade

2013-17 Anders Bruun Jonassen: Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance

2014-1 David Sloth Pedersen: A Journey into the Dark Arts of Quantitative Finance

2014-2 Martin Schultz-Nielsen: Optimal Corporate Investments and Capital Structure

2014-3 Lukas Bach: Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods

2014-4 Tanja Groth: Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomic analysis

2014-5 Niels Strange Hansen: Forecasting Based on Unobserved Variables

2014-6 Ritwik Banerjee: Economics of Misbehavior

2014-7 Christina Annette Gravert: Giving and Taking – Essays in Experimental Economics

2014-8 Astrid Hanghøj: Papers in purchasing and supply management: A capability-based perspective

2014-9 Nima Nonejad: Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics

2014-10 Tine L. Mundbjerg Eriksen: Essays on Bullying: an Economist's Perspective

2014-11 Sashka Dimova: Essays on Job Search Assistance

2014-12 Rasmus Tangsgaard Varneskov: Econometric Analysis of Volatility in Financial Additive Noise Models

2015-1 Anne Floor Brix: Estimation of Continuous Time Models Driven by Lévy Processes

2015-2 Kasper Vinther Olesen: Realizing Conditional Distributions and Coherence Across Financial Asset Classes

2015-3 Manuel Sebastian Lukas: Estimation and Model Specification for Econometric Forecasting

2015-4 Sofie Theilade Nyland Brodersen: Essays on Job Search Assistance and Labor Market Outcomes

2015-5 Jesper Nydam Wulff: Empirical Research in Foreign Market Entry Mode

2015-6 Sanni Nørgaard Breining: The Sibling Relationship Dynamics and Spillovers

2015-7 Marie Herly: Empirical Studies of Earnings Quality

2015-8 Stine Ludvig Bech: The Relationship between Caseworkers and Unemployed Workers

2015-9 Kaleb Girma Abreha: Empirical Essays on Heterogeneous Firms and International Trade

2015-10 Jeanne Andersen: Modelling and Optimisation of Renewable Energy Systems

2015-11 Rasmus Landersø: Essays in the Economics of Crime

2015-12 Juan Carlos Parra-Alvarez: Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)

2015-13 Sakshi Girdhar: The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis

2015-14 Wenjing Wang: Corporate Innovation, R&D Personnel and External Knowledge Utilization

2015-15 Lene Gilje Justesen: Empirical Banking

2015-16 Jonas Maibom: Structural and Empirical Analysis of the Labour Market

2015-17 Sylvanus Kwaku Afesorgbor: Essays on International Economics and Development

2015-18 Orimar Sauri: Lévy Semistationary Models with Applications in Energy Markets

2015-19 Kristine Vasiljeva: Essays on Immigration in a Generous Welfare State

2015-20 Jonas Nygaard Eriksen: Business Cycles and Expected Returns

2015-21 Simon Juul Hviid: Dynamic Models of the Housing Market

2016-1 Silvia Migali: Essays on International Migration: Institutions, Skill Recognition, and the Welfare State

2016-2 Lorenzo Boldrini: Essays on Forecasting with Linear State-Space Systems

2016-3 Palle Sørensen: Financial Frictions, Price Rigidities, and the Business Cycle

2016-4 Camilla Pisani: Volatility and Correlation in Financial Markets: Theoretical Developments and Numerical Analysis

2016-5 Anders Kronborg: Methods and Applications to DSGE Models

2016-6 Morten Visby Krægpøth: Empirical Studies in Economics of Education

2016-7 Anne Odile Peschel: Essays on Implicit Information Processing at the Point of Sale: Evidence from Experiments and Scanner Data Analysis

2016-8 Girum Dagnachew Abate: Essays in Spatial Econometrics

2016-9 Kai Rehwald: Essays in Public Policy Evaluation

2016-10 Reza Pourmoayed: Optimization Methods in a Stochastic Production Environment

2016-11 Sune Lauth Gadegaard: Discrete Location Problems – Theory, Algorithms, and Extensions to Multiple Objectives

2016-12 Lisbeth Palmhøj Nielsen: Empirical Essays on Child Achievement, Maternal Employment, Parental Leave, and Geographic Mobility

2016-13 Louise Voldby Beuchert-Pedersen: School Resources and Student Achievement: Evidence From Social and Natural Experiments

2016-14 Mette Trier Damgaard: Essays in Applied Behavioral Economics

2016-15 Andrea Barletta: Consistent Modeling and Efficient Pricing of Volatility Derivatives

2016-16 Thorvardur Tjörvi Ólafsson: Macrofinancial Linkages and Crises in Small Open Economies

2016-17 Carlos Vladimir Rodríguez Caballero: On Factor Analysis with Long-Range Dependence

2016-18 J. Eduardo Vera-Valdés: Essays in Long Memory