Essays in Volatility Modelling and Applied Econometrics

Jose Angel Alcantara Lizarraga
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Economics Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/2389

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Economics

Dissertation Examination Committee:
Werner Ploberger, Chair
Gaetano Antinolfi
Ian Fillmore
Jose Figueroa Lopez
Robert Parks

Essays in Volatility Modelling and Applied Econometrics
by
Jose Angel Alcantara Lizarraga

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2021
St. Louis, Missouri
Contents

List of Figures ........................................ iv
List of Tables ........................................ vi
Acknowledgements .................................... viii
Abstract ............................................. ix

1 Forecasting Realized Variance: Long Memory, Semivariances
   and Signed Jumps ................................. 1
   1.1 Introduction and Literature ...................... 1
   1.2 Realized Framework .............................. 6
       1.2.1 Benchmark Model: The HAR Model .......... 8
       1.2.2 The Realized Semivariances and Signed Jumps . 9
   1.3 Modelling volatility with Long Memory, RS and Signed Jumps ......... 10
   1.4 Simulation ..................................... 13
   1.5 Empirical exercise .............................. 20
       1.5.1 Data and in-sample results ............... 20
       1.5.2 Out-of-Sample Forecasts ................. 22
       1.5.3 Error Diagnostic .......................... 23
   1.6 Conclusions ................................... 23
   1.7 Tables and figures ............................. 26
       1.7.1 Figures .................................. 26
       1.7.2 In-Sample Statistics ...................... 35
       1.7.3 Error Diagnostic ........................ 38

2 Forecasting Volatility: An empirical R-GARCH Approach for the Mexican IPC 44
   2.1 Introduction ................................... 44
2.2 Econometric Methodology ........................................... 47
  2.2.1 Realized GARCH ................................................... 47
  2.2.2 The Model ....................................................... 48
  2.2.3 Data ............................................................ 51
  2.2.4 Estimation ....................................................... 55
  2.2.5 In-Sample Results ............................................. 55
  2.2.6 Out-of-Sample Results ....................................... 59
  2.2.7 Conclusions .................................................... 59

3 Income Inequality and free trade: An applied causality analysis for Mexico 61
  3.1 Introduction ......................................................... 61
  3.2 Literature Review .................................................. 65
  3.3 Methodology and data ............................................ 67
  3.4 Results ............................................................ 70
  3.5 Conclusion ........................................................ 74
  3.6 Statistical Results ................................................ 76
    3.6.1 Unit Root Tests ............................................... 76
    3.6.2 Impulse Response Functions ................................. 77

Bibliography ........................................................... 79
List of Figures

1.1 Simulation Exercise .................................. 15
1.2 Simulation Signed Jumps ............................... 16
1.3 59 % of the time the $RV_m > RV_d$ ................... 16
1.4 Euro­Index and Realized Variances (2014­2020) .......... 26
1.5 Euro­Index ACF of Realized Variances ................. 27
1.6 Hong Kong­Index and Realized Variances (2014­2020) .. 28
1.7 Hong Kong­Index ACF of Realized Variances .......... 29
1.8 USA­Index and Realized Variances (2014­2020) ......... 30
1.9 USA­Index ACF of Realized Variances ................. 31
1.10 $\Delta J^2 = (RSV^+ - RSV^-)/RV_t$ .................. 32
1.11 Signed Jumps $\Delta J$ for the 3 indexes ................. 33
1.12 Residuals USA INDEX Models .......................... 39
1.13 Residuals EURO INDEX Models .......................... 40
1.14 Residuals HK INDEX Models ............................ 41
1.15 Out­Of­Sample Results ................................. 42

2.1 Mexican IPC­ RV and ACF (2000­2017) ................. 52
2.2 Mexican IPC­ $S_t$ and Asymmetry .................... 54

3.2 Annual average increase income (CEPAL) .................. 65
3.4 Impulse Response functions Trade/GDP and Labor Income Gini ............... 77
3.5 Impulse Response functions Trade/GDP and Hourly Wage Gini ............. 77
3.6 Impulse Response functions Trade/GDP and Total Disposable Income Gini (SWIID) 78
List of Tables

1.1 Simulation Forecasting Performance ........................................ 17
1.2 Forecasting Relative Gains .................................................... 18
1.3 Forecasting Relative Gains II .................................................. 19
1.4 Summary Statistics for log intraday Returns ............................... 33
1.5 Summary Statistics for RV ..................................................... 33
1.6 Summary Statistics for $RV_m$ ................................................ 34
1.7 Summary Statistics for $|RV_d - RV_m|$ ...................................... 34
1.8 Summary Statistics for $(RV_d - RV_m)^+$ .................................. 34
1.9 Summary Statistics for $(RV_d - RV_m)^-$ .................................. 34
1.10 (HAR) $RV_{d,t} = \beta + \beta_d RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ ......................................................... 35
1.11 In Sample.(HAR-Q) $RV_{d,t} = \beta + (\gamma + \beta Q_{t-1}^{1/2}) RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ ............................................... 35
1.12 In Sample.(TV-HAR) $RV_{d,t} = \beta + (\gamma + \alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ ................................................. 36
1.13 In Sample.(SV-HAR) $RV_{d,t} = \beta + R_{d,t-1}^+ + \beta^- R_{d,t-1}^- + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ .................................................. 36
1.14 In Sample.(TV-HAR) $RV_{d,t} = \beta + (\alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \beta^+ R_{d,t-1}^+ + \beta^- R_{d,t-1}^- + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ ............................................ 36
1.15 In Sample.(SV-TVHAR) $RV_{d,t} = \mu + \phi D_{d,t-1} + \phi^C B_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ .................................................. 37
1.16 In Sample.(SJTV-HAR) $RV_{d,t} = \mu + \phi D_{d,t-1} + \phi^C B_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ .................................................. 37
1.17 In Sample.(SJ-HAR) $RV_{d,t} = \mu + \phi D_{d,t-1} + \phi^C B_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ .................................................. 37
1.18 In Sample.(SSJ-HAR) $RV_{d,t} = \mu + \phi D_{d,t-1} + \phi^C B_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$ .................................................. 37
1.19 In Sample (SSJ-TVHAR) \( RV_{d,t} = \mu + \alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \phi^J \Delta J_{t-1}^2 + \phi^J - \Delta J_{t-1}^2 + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t \) .................................................. 37

1.20 Errors Diagnostic (SV-TVHAR) \( RV_{d,t} = \beta + (\alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \beta^+ RS^+_t + \beta^- RS^-_t + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t \) .................................................. 38

1.21 Errors Diagnostic (SJTV-HAR) \( RV_{d,t} = \mu + \alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \phi^J \Delta J_{t-1} + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t \) .................................................. 38

1.22 Errors Diagnostic (SSJ-TVHAR) \( RV_{d,t} = \mu + \alpha |RV_{d,t-1} - RV_{m,t-1}|) RV_{d,t-1} + \phi^J + \Delta J_{t-1}^2 + \phi^J - \Delta J_{t-1}^2 + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t \) .................................................. 38

1.23 Relative Gains MSE Euro-Index .................................................. 43

1.24 Relative Gains MSE HK-Index .................................................. 43

1.25 Relative Gains MSE USA-Index .................................................. 43

2.1 Summary Statistics ................................................................. 53

2.2 In-Sample Analysis IPC ............................................................ 58

2.3 LR-TEST ................................................................. 58

2.4 Out-of-Sample Analysis IPC ............................................................ 59
Acknowledgments

I would first like to thank my advisor, Professor Werner Ploberger, whose teaching and advise were invaluable in formulating the research questions of my dissertation. He made me discover new insights of a fascinating field. I also thank Bob and Jose Figueroa for their invaluable support and advise. Finally, I would like to thank my parents and my family for always believe in me and for always stay there in the most difficult moments.

Jose Angel Alcantara Lizarraga

Washington University in St. Louis

May 2021
Chapter 1 examines the impact of signed realized measures on forecasting realized volatility using a Time Varying Heterogeneous Autoregressive Model (HAR) framework. To account for this we propose an empirical methodology contribution considering a family of 3 extended models (Signed TV-HAR), where the disentangled continuous and jump signed variations are incorporated. We demonstrate empirically (using 3 major financial indexes for the period 2014-2020) and via Monte Carlo simulations that the proposed family produces significantly better forecasts than the standard Corsi (2009)’s HAR model and some of its classical extensions. The purpose of chapter 2 is to estimate and forecast the Mexican IPC (Quotes and Trade) Index. We introduce a new model that extends the Realized GARCH models of Hansen et al. (2012). Our model generalizes the original specification along three different directions. First, I adopt an autoregressive specification for the volatility dynamics. Second, it features a time varying volatility persistence. That is, the response coefficient in the volatility equation is time sensitive. Finally, our framework allows to consider, in a parsimonious way, the inclusion of a jump
adjusted realized measure in the measurement equation of the model. The forecasting performance of the model is evaluated obtaining gains relative to the traditional R-GARCH model. In chapter 3 the objective is to establish if there are Granger causality relationships between income inequality and trade in Mexico. What is the statistical relationship between globalization and income inequality in Mexico? Is there a causal relationship between inequality and trade volume in Mexico?
Chapter 1

Forecasting Realized Variance: Long Memory, Semivariances and Signed Jumps

1.1 Introduction and Literature

Volatility of asset returns has been one of the primary objects of study in modern financial econometrics for the recent 30 years. The reason is that volatility forecast is crucial for many investment decisions primarily because they are relevant for option pricing, asset allocation, risk management and public policy decision making. Hence, accurately measuring and forecasting financial volatility is of crucial importance for financial market participants.

The non-parametric approach has become the new standard in volatility forecasting. Ying (1996) and Clark (1973) were the first to use absolute price changes as a proxy of volatility. Later, Poterba and Summers (1984), French et al. (1987), Pagan and Schwert (1990) used ex-post sample variances computed from higher frequencies returns data as lower frequency volatility measures.
More recently, with the availability of High Frequency (HF) data, the field has seen a significant improvement in the development of new models. Now, we can compute daily variances from intraday returns. It is now known that returns are weak signals of the level of volatility making conventional GARCH models not suited for modeling situations in which the volatility jumps abruptly. Hence, incorporating realized measures based on High Frequency (HF) data can improve the estimation procedure and fix this shortcoming (Hansen and Huang, 2016).

The fundamental idea is that accurate estimates of volatility are essential prerequisites for good forecasts. Hence, the development of high frequency estimators (realized measures) has contributed to the progress of this field by improving our knowledge of financial volatility and its dynamic properties (See Barndorff-Nielsen, et.al 2002). In particular, construction of observable Realized Variance ($RV$) series from high frequency financial market data is standard practice in finance. That is due to the development of modern mathematical and statistical tools, the increasing availability of tick data and in general the technological advancement that has improved the trading strategies in financial markets (Ait-Sahalia, 2014).

Asymmetries and persistence (slowly decaying auto-correlation) in the DGP for asset returns are known stylized facts in the High Frequency Literature.

In particular, long memory or persistence is one of the most studied time series features (or stylized facts) of financial returns (Ding et al. 1993; Bollerslev and Ole Mikkelsen 1996; Breidt et al 1998; Yalama and Elik 2013). Greene and Fielitz (1977) were the first to observe this pattern in financial data. Some of the historical developments are discussed by Baillie (1996). It refers to the fact that the autocorrelation of the squared or absolute returns of financial assets, as a proxy for underlying volatilities, decay at a slow rate. Different models that account for this include Baillie et al. (1996) who proposed the fractionally integrated GARCH model (FIGARCH) to account for
long memory volatility and Andersen et al. (2003) used this framework to model Realized Variance (RV). The Markov Switching Model in Maheu and McCurdy (2002); The Unobserved Component models of Barndor-Nielsen and Shephard (2002a) and Koopman, Jungbacker, and Hol (2005); and the Mixed Data Sampling (MIDAS) approach of Ghysels et al. (2006).

One of the most important examples of the long memory feature is found in the Realized Volatility ($RV$). Under suitable conditions Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2001, 2002a) found that the Realized Volatility ($RV$), defined as the sum of the intra-day returns, is an unbiased and efficient estimator of the return volatility. $RV$ has become a milestone in the analysis of volatility forecasting due to the availability of High Frequency Data. This paper focuses on modeling and forecasting $RV$.

The usage of $RV$ in empirical finance is a common practice since it reduces the need to formulate a particular model and brings a parsimonious direct measurement of the volatility. Many authors have tried to explain the long memory feature of the $RV$, hence there are some extensions of the standard models using (HF) data which are now the foundations of the growing research in volatility forecasting in the context of long persistence. Engle (2002) was the first to introduce the realized volatility as an explicit regressor in a single equation model of the GARCH type. The goodness of fit improved significantly as the AR component of the model became insignificant.

Corsi (2009) introduced the Heterogenous Auto Regressive (HAR) model by specifying the realized volatility within a constrained autoregressive model. The author considers a genuine long memory data generating process by making use of Lebanon’s (2001) idea that a true long process is the result of the aggregation of only a few heterogeneous time scales. Most recent literature makes use of the HAR model and its extensions for forecasting RV and it is arguably nowadays the benchmark model in the field.
Many extensions to the model have been accomplished. Shephard and Sheppard (2010) developed the HEAVY model as an extension of the HAR. Then and Ghysels (2011) use the HAR model to study the impact of news in volatility. Andersen et al. (2011) extends the model considering the presence of jumps in order to describe the dynamic dependencies in the daily continuous sample path variability. In a more recent work, Liu et al. (2015) studied the accuracy of a variety of estimators of asset price variation using HF data. The other class of extensions are related to the relaxation of the constant coefficients assumption in the model.

Two extensions are of particular relevance for this paper. First, Patton and Sheppard (2011) examine the impact of Realized Semivariances and Signed Jumps on future volatility. Signed measures of volatility are relatively new concepts incorporated in the High frequency literature that are related to positive and negative returns independently. They find for 150 stocks that incorporating those measures in the HAR model improves volatility forecasts. The authors point out that although high frequency returns of financial assets are small, their signs are important and have a direct impact on volatility forecasting.

The second work that is of particular importance for this paper is Li’s (2016). The latter extends the traditional HAR model by considering a time varying specification of the coefficient of the first lag of the model. The author points out that the difference between long term moving average of the daily RV, which is the $RV_m$ and $RV_d$, that is $(RV_d - RV_m)$ is negative with more frequency than when the difference is positive. This creates an scenario in which RV tends to quickly revert back when there is an increase in lag $RV_d$ compared with its longer-term average level, and it tends to lower revert back when lag $RV_d$ decreased compared to $RV_m$.

To the best of my knowledge there is no analysis in the literature that studies the HAR specification in the context of time varying coefficients and accounting for the presence of jumps and
signed measures effects in terms of forecasting performance.

This paper contributes to this literature in two directions and aims to fill this apparent gap. In the first place, I disentangle the jump variation and the continuous variation of the process using the recently proposed concepts of Realized Semi variance and Signed Jumps. In the second place, I propose a simple model specification that combines the latter effects with the flexibility that the coefficient of the first lagged of the daily Realized Variance is time varying. Our specification is extremely simple to implement since it is based in the parsimonious HAR structure.

This paper proposes a methodology contribution to forecast Realized Volatility using High Frequency Data accounting for persistence by using a time varying specification in the Corsi’s (2009) HAR model. In section 2, I introduce briefly the realized framework for analyzing High Frequency data; in section 3, I describe the proposed model; in section 4,a Montecarlo Simulation is designed to test the forecasting performance of the proposed models; in section 5 an empirical exercise is performed using 3 financial indexes. Finally, section 5 concludes with the final remarks and suggested future research questions.
1.2 Realized Framework

In the last few years, the availability of high frequency financial data has allowed researchers to build accurate measures of the latent daily volatility based on intraday returns. By allowing a continuous time specification of the price process we can set up a theoretical background of the Realized Measures. For day $t = 1, 2, 3..., $ denote the intra-day price as $P_t$. Let $p_t = \log(P_t)$ be determined by the following process:

$$dp_t = \mu_t dt + \sigma_t dW_t + \kappa_t dN_t,$$

for $0 \leq t \leq T$, where $\mu_t$ and $\sigma_t$ are the drift and instantaneous volatility processes, respectively, $W_t$ is a standard Brownian Motion, and $N_t$ is a finite activity counting process. Under the conditions of the absence of Microstructure Noise and $dN_t = 0$, $p_t$ follows a continuous semi-martingale process.

Then, the daily log return is:

$$R_t = \log(P_t) - \log(P_{t-1})$$

The Quadratic Variation ($QV$) of those log-returns coincides with the Integrated Variance:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds$$
A constant daily return variance estimator is naturally:

$$V = \frac{1}{n} \sum_{t=1}^{n} R_t^2$$  \hspace{1cm} (1.4)$$

Each day $t$ can be divided into $N$ equidistant intervals of size $\Delta$. The log closing price at the $i$-th interval of day $t$ is denoted $P_{t-1+i\Delta}$. The intra-day returns for the $i$-th $\Delta$ period are then defined as:

$$r_{t,i} = \log(P_{t-1+i\Delta}) - \log(P_{t-1+(i-1)\Delta}),$$  \hspace{1cm} (1.5)$$

for $i = 0, \ldots, N$. The Realized Variance (RV) is simply the sum of the $N$ intra day squared returns, at frequency $\Delta$:

$$RV_t^\Delta = \sum_{i=1}^{N} r_{t,i}^2$$  \hspace{1cm} (1.6)$$

The IV and the conditional variance do not coincide, but as shown by Andersen et al.(2001), we have that:

$$E(IV_t | F_{t-1}) = V(r_t | F_{t-1}),$$  \hspace{1cm} (1.7)$$

where $V(r_t | F_{t-1})$ is the conditional variance of the returns. The RV is a consistent estimator of the true latent volatility when $\Delta \to 0$ and $dN_t = 0$. When $dN_t \neq 0$, this convergence is no longer true. Let $dN_t = \kappa_t dq_t$, where $\kappa_t$ is a measure of the size of the jump in prices and $q_t$ is a counting process. Then we have:

$$RV_t \to \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1 \leq s \leq t} \kappa_s^2(s)$$  \hspace{1cm} (1.8)$$
1.2.1 Benchmark Model: The HAR Model

The model is based on two moving averages of the daily RV, namely the weekly RV ($RV_w$) and the monthly RV ($RV_m$) to model the long memory behavior of volatility. The weekly and monthly moving averages are defined at time $t$, respectively, as:

$$RV_{w,t} = \frac{1}{5}(RV_t + RV_{t-1} + \ldots + RV_{t-4}),$$

(1.9)

$$RV_{m,t} = \frac{1}{22}(RV_t + RV_{t-1} + \ldots + RV_{t-21}).$$

(1.10)

Based on the latter definitions, the HAR model (Corsi, 2009) is defined as:

$$RV_{d,t} = \beta + \beta_d RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t,$$

(1.11)

where the term $u_t$ is a zero mean innovation process. The model is parsimoniously estimated by standard OLS techniques. Following the success of the HAR model, many extensions have been proposed considering jumps, leverage effects and nonlinearities. Corsi, Pirino and Reno (2010) used the C-Tz test to detect jumps and include them in the model. Bollerslev, Litvinova, and Tauchen (2006) considered the presence of asymmetric leverage effects to forecast the volatility. The inclusion of structural breaks and regime-switching is considered in McAleer and Medeiros (2008) and Scharth and Medeiros (2009). Patton and Shepard (2015) extended the model considering the Realized Semivariances showing that future volatility is more influenced by the volatility of past negative returns.

The other class of extensions of the model is related to the possibility that the coefficients
are not constant over time. Perhaps the most known model in this category is the HAR-Q model
developed by Bollerslev, et al. (2016). The foundation of the model is that RV is equal to the
sum of two components: the true latent IV and a time varying measurement error. It allows the
autocorrelation parameter to vary with the estimated degree of measurement error.

More recently, Li (2016) proposed a new type of Time Varying Model based on the idea that
since the monthly RV is a moving average of the daily RV, the number of observations of daily RV
that are higher than the monthly moving average should be equal to the number of observations
of daily RV that are lower than its monthly moving average. Using evidence from some SP-500
financial stocks the pattern works in the opposite direction. In the empirical application presented
at the end of this paper, I present evidence that it is actually the case for 3 exchange rates and 3
financial indexes selected for analysis. As a way to overcome this feature of the data that causes
less forecasting power of the HAR and HARQ model (see Li’s work (2016)), the HAR model is
extended in its TV-HAR specification:

\[
RV_{d,t} = \beta + (\gamma + \alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \beta_wRV_{w,t-1} + \beta_mRV_{m,t-1} + u_t. \tag{1.12}
\]

1.2.2 The Realized Semivariances and Signed Jumps

To decompose volatility into components that relates only to positive and negative high frequency
returns, respectively, I use the Realized Semivariance \( RS_t^+ \) \( RS_t^- \) proposed by Barndorff-Nielsen
and Shephard (2007):

\[
RS_t^+ = \sum_{i=1}^{m} r_{t,i}^2 1_{\{r_{t,i} > 0\}} \rightarrow \frac{1}{2} \int_{t-1}^{t} \sigma_s^2 ds + \sum_{t-1 < s \leq t} \Delta p_s^2 1_{\{\Delta p_s > 0\}}, \tag{1.13}
\]
where $1_{\{\cdot\}}$ is an indicator function and $\Delta p_s = p_s - p_{s-}$ captures a jump, if it present. Note that $RV_t = RS_t^+ + RS_t^-$. 

According to Patton and Sheppard (2015), we can define the signed jump variation as:

$$\Delta J^2_t = RS_t^+ - RS_t^- \rightarrow \sum_{t-1<s\leq t} \Delta p_{s}^2 1_{\{\Delta p_s>0\}} - \sum_{t-1<s\leq t} \Delta p_{s}^2 1_{\{\Delta p_s<0\}}$$  \hspace{1cm} (1.15)$$

The jump variation can be decomposed in the following way:

$$\Delta J^2_t = \Delta J^2_{t}^{+} + \Delta J^2_{t}^{-} = (RS_t^+ - RS_t^-)1_{\{(RS_t^+ - RS_t^-)>0\}} + (RS_t^+ - RS_t^-)1_{\{(RS_t^+ - RS_t^-)<0\}}.$$  \hspace{1cm} (1.16)$$

### 1.3 Modelling volatility with Long Memory, $RS$ and Signed Jumps

Given the measures described above it is our objective to study their forecasting power for future realized volatility. To that end I propose a model based on the HAR framework and extended by considering Realized Semi-Variances and a time varying specification. To the best of my knowledge this is the first empirical work that considers to extend the HAR model with signed realized measures and consider a time varying coefficient specification at the same time. The realized volatility features long memory. The HAR model, although it does not formally belong to the class of long memory models, it deploys a parsimonious approximation which is able to accurately
simulate this stylized fact.

As we have seen, several extensions based on the HAR model have been proposed in the recent literature: Andersen et al. (2007) and Corsi et al. (2012) include a jump component in the regression equation. More recently, Patton, A.J. and Sheppard, K and Chen, X and Ghysels, E (2010) considered the use of semi variances to forecast volatility. Audrino el al. (2016) extended the model by considering jumps in the volatility and in the returns in the regression. None of the previous studies, to the best of my knowledge, used the realized semi variances and signed jumps in the context of a time varying specification to generalize the HAR model.

The empirical evidence shows that jumps have an impact on future volatility. Their effect can be interpreted as a leverage effect. Disentangling jump variation and continuous variation and downside semi variation and upside semi variation of the Realized Variance improves volatility forecasts. I combine the ideas of Patton et al (2010) and Li (2015) by considering the following family of time varying signed models:

**MODEL 1**

The Signed Jumps Time Varying HAR (SJ-TVHAR)

\[
RV_{d,t} = \mu + (\gamma + \alpha |RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \phi^J \Delta J_t^2 + \phi^C BV_t + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t
\] (1.17)
MODEL 2

The Semi-Variance Time Varying HAR (SV-TVHAR)

\[
RV_{d,t} = \beta + (\gamma + \alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1}
\]

\[+\beta(+)RS_t^+ + \beta(-)RS_t^-\]

\text{Semi variance Decomposition}

\[+\beta w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t\]

(1.18)

MODEL 3

The Split Signed Jumps Time Varying HAR (SSJ-TVHAR)

\[
RV_{d,t} = \mu + (\gamma + \alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1}
\]

\[+\phi^+ \Delta J_t^2 + \phi^- \Delta J_t^2 + \phi^C \text{BV}_t + \beta w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t\]

(1.19)

In models 1 and 3, the term \(\text{BV}_t\) is the "Bipower Variation", an estimator presented by Barndorff-Nielsen and Shephard (2004) who showed that:

\[
\text{BV}_t = \frac{\pi}{2} \sum_{i=2}^{m} |r_{t,i}||r_{t,i-1}| \rightarrow \int_{t-1}^{t} \sigma_s^2 ds \quad m \rightarrow \infty
\]

(1.20)

That is, \(\text{BV}_t\) is a jump robust estimator.

At first sight, these models are parsimonious, that is, they are not over parametrized. They consist of 7 parameters easily estimated by standard linear regression techniques. The models are estimated by OLS.
These specifications will allow us to study the effect of signed jumps in the forecasting power of the models. It is well known in the literature that jumps in the price process affects the forecasting performance of the HAR model. The basic extension of the HAR model with jumps (HAR-J) considers to measure the jump in a particular day as the difference:

$$\max\{RV_t - BV_t, 0\} \quad (1.21)$$

This, however, implies a loss in the information provided by the nature of the jumps (that is, if they are associated with negative or positive returns). Considering semi variances in the model will account for such stylized fact.

### 1.4 Simulation

This section presents a Monte Carlo simulation study to investigate the performance of the proposed model. The simulations are based in the following stochastic volatility model:

$$dp(t) = \eta dt + g(t) dW_p(t) + dZ(t), \quad (1.22)$$

$$g^2(t) = exp\{\gamma_1 v(t)\}, \quad (1.23)$$
\[ dv(t) = \alpha v(t) dt + dW_v(t), \quad (1.24) \]

where \( Z_t \) is a jump process. A compound Poisson process with intensity \( \lambda \) is selected where the jump size follows a standard normal distribution, that is:

\[ Z_t = \sum_{i=1}^{N(t)} Y_i; t \geq 0, \quad (1.25) \]

where \( N(t) \) is a compound process with intensity \( \lambda \), \( Y_i \) is a sequence of \( i.i.d \) variables. \( W_{p,t} \) and \( W_{v,t} \) are 2 standard and correlated Brownian Motions with \( \rho = -0.62 \). The element \( g(t) \) is a persistent parameter in the model in order to generate high autocorrelations. This model has been used by Huang and Tauchen (2005). The price drift is set \( \eta = 0.03 \), while for the volatility factor we have \( \alpha = 0.1 \) and \( \gamma_1 = 0.125 \). I simulate data for the unit interval \([0, 1]\) and normalize 1 minute to be \( 1/390 \), \( (\Delta t = 1/390) \). Simulations are generated using an Euler Scheme and 1-minute data is obtained considering \( T = 730 \) days. Then, the 1-factor volatility model generates \( 390 \times 730 = 284,700 \) price observations. Finally, I compute the Realized Variance.
Figure 1 shows a path of simulated prices and returns along with the computed simulated realized variance and its Autocorrelation function. It mimics a scenario of long memory of the volatility.

Table 1 shows the result of the forecasting performances comparison of the HAR, HARQ, and the TV-HAR Models proposed in this paper. The comparison is based on the MSE and MAE. I consider sequences of one-day-ahead rolling forecasting, with a rolling window of 365 days and 183 days (six months). Table 2 presents the relative gains of the models relative to the set of models selected for the comparison.

Both the SJ-TVHAR and the SV-TVHAR models outperform the HAR, HARQ and TV-HAR
models using the MSE and MAE loss functions and for the two different forecasting horizon considered. In the case of the SJ-TVHAR, it produces better forecasts than the model with Signed Jumps with no time varying coefficient in Patton et al. (2015) SJ. It also produces marginal better forecasts than the SV model. In the case of the SV-TVHAR model, according to the simulation, it has better performance than the SV model, but very similar performance relative to the SJ model. The model that better performs in general is the SSJ-TVHAR, beating the HAR, HARQ, TV-HAR and SV-TVHAR, for the two out-of-sample windows considered.

\[(a) \% \Delta J^2 = (RSV^+ - RSV^-)/RV_i \]

Figure 1.2: Simulation Signed Jumps

\[\text{Figure 1.3: 59% of the time the } RV_m > RV_d\]

16
Table 1.1: Simulation Forecasting Performance

<table>
<thead>
<tr>
<th>Model Type</th>
<th>MSE 12 months</th>
<th>MAE 12 months</th>
<th>MSE 6 months</th>
<th>MAE 6 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAR</td>
<td>32.11648</td>
<td>21.56833</td>
<td>43.17789</td>
<td>32.04452</td>
</tr>
<tr>
<td>HARQ</td>
<td>32.02582</td>
<td>21.50840</td>
<td>43.04456</td>
<td>31.82687</td>
</tr>
<tr>
<td>TV-HAR</td>
<td>31.93050</td>
<td>21.45810</td>
<td>42.75232</td>
<td>31.55790</td>
</tr>
<tr>
<td>SJ</td>
<td>30.05155</td>
<td>20.37335</td>
<td>40.66162</td>
<td>30.14317</td>
</tr>
<tr>
<td>SJ-TVHAR</td>
<td>29.98443</td>
<td>20.34344</td>
<td>40.54683</td>
<td>30.04503</td>
</tr>
<tr>
<td>SV-HAR</td>
<td>30.39809</td>
<td>20.16763</td>
<td>41.46368</td>
<td>29.73728</td>
</tr>
<tr>
<td>SV-TVHAR</td>
<td>30.28269</td>
<td>20.21975</td>
<td>41.14904</td>
<td>29.64475</td>
</tr>
<tr>
<td>SSJ-HAR</td>
<td>30.05037</td>
<td>20.36974</td>
<td>39.46797</td>
<td>39.46797</td>
</tr>
<tr>
<td>SSJ-TVHAR</td>
<td>29.96576</td>
<td>20.33294</td>
<td>39.34046</td>
<td>29.47102</td>
</tr>
<tr>
<td>HAR-J</td>
<td>32.09233</td>
<td>21.59609</td>
<td>47.91232</td>
<td>41.97823</td>
</tr>
</tbody>
</table>

The forecasting performance is based on the $MSE$ function of the Realized variance Forecasts. $MSE = \frac{1}{n}(RV_{t,t+h} - RV)^2$. 12 and 6 months refer to the forecasting out-of-sample interval to be forecasted.
### Table 1.2: Forecasting Relative Gains

<table>
<thead>
<tr>
<th></th>
<th>HAR</th>
<th>HARQ</th>
<th>TV-HAR</th>
<th>SJ</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.0664</td>
<td>0.06374</td>
<td>0.06094</td>
<td>0.06712</td>
<td>0.01361</td>
</tr>
<tr>
<td>6 months</td>
<td>0.061124</td>
<td>0.058027</td>
<td>0.05159</td>
<td>0.00282</td>
<td>0.01452</td>
</tr>
<tr>
<td><strong>MAE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.0568</td>
<td>0.054163</td>
<td>0.051945</td>
<td>0.001468</td>
<td>-0.0087174</td>
</tr>
<tr>
<td>6 months</td>
<td>0.062395</td>
<td>0.05598</td>
<td>0.047939</td>
<td>0.003255</td>
<td>-0.01034</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.0571</td>
<td>0.05442</td>
<td>0.05160</td>
<td>-0.00769</td>
<td>0.0037962</td>
</tr>
<tr>
<td>6 months</td>
<td>0.0469</td>
<td>0.04403</td>
<td>0.037501</td>
<td>-0.01198</td>
<td>0.00758</td>
</tr>
<tr>
<td><strong>MAE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.0625</td>
<td>0.0599</td>
<td>0.0577</td>
<td>0.00753</td>
<td>-0.00258</td>
</tr>
<tr>
<td>6 months</td>
<td>0.07488</td>
<td>0.0656</td>
<td>0.06062</td>
<td>0.0165</td>
<td>0.00311</td>
</tr>
</tbody>
</table>

MAE and MSE means the ratio of the losses for the different models relative to the losses of the HAR model. The MAE and MSE Gains(%) measure the gains of forecasting accuracy compared with the HAR model, for example, the MAE gains of HARQ is calculated as follows: \((\text{MAE}_{\text{HAR}} - \text{MAE}_{\text{HARQ}})/\text{MAE}_{\text{HAR}}\).
Table 1.3: Forecasting Relative Gains II

<table>
<thead>
<tr>
<th></th>
<th>SJ-TVHAR</th>
<th>SSJ-TVHAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SV-TVHAR</td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.009849</td>
<td>-0.000623</td>
</tr>
<tr>
<td>6 months</td>
<td>0.014634</td>
<td>-0.0306</td>
</tr>
<tr>
<td>MAE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>-0.00611</td>
<td>-0.000516</td>
</tr>
<tr>
<td>6 months</td>
<td>-0.0135</td>
<td>-0.0194</td>
</tr>
<tr>
<td>SV-TVHAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>-0.010576</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>0.00556</td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 months</td>
<td>0.00556</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>-0.00589</td>
<td></td>
</tr>
</tbody>
</table>
1.5 Empirical exercise

1.5.1 Data and in-sample results

I implement the proposed family of TV-HAR models in the case of the 3 major financial indexes and discuss its implications in terms of forecasting performance relative to the traditional HAR model and its traditional extensions, that is versus the Andersen’s HAR-J model and the Bollerslev’s HARQ model.

Intra-daily 5 minute data have been extracted from Dukascopy for the period 00:00:00 GMT, June 2, 2014 through 23:55:00 GMT, May 20, 2020 (627,840 observations). We divide the sample into 3 in-sample period (3 months, 12 months and 24 months) with their corresponding out-of-sample periods.

The empirical analysis relies on 3 financial indexes: The Euro Index 50, the Hong Kong Index, and the USA 500 index. Since trading is slower over weekends, I follow the adjustment process of Andersen and Bollerslev (1998) by excluding the returns from Friday 21:00 GMT to Sunday 21:00 GMT. Stamps with volume zero are also discarded. Summary statistics of the data can be found at the end of the paper. Data outside the interval of time 09:30:00 to 16:00:00 are also not considered in the analysis. I make use of realized variance computations using 5-minute data, since it is generally accepted in the literature that this sampling frequency is the most robust to microstructure noise.

We can note that the 3 series exhibit long memory in their Realized Variance. I report the number of observations, mean, standard deviation, minimum and maximum values for daily, weekly and monthly RVs. Figures 2-7 show the ACF for the Realized Variance measures. Tables 14 and 15 show the summary statistics for \((RV_d - RV_m)^+\) and \((RV_d - RV_m)^-\). There is evidence that on average, 35% of the difference is positive, while 65% is negative. This is in line with the results of
Li (2016). The standard deviations of the daily RV are higher than those of their moving averages $RV_w$ and $RV_m$. We can see a different pattern for the daily RV reverting back to its long term average level.

In the empirical analysis below I use an interesting consequence of equation (15). That is, the variation due to the continuous component can be removed by subtracting $RS^-$ from $RS^+$, giving as a result the **Signed Jump Variation** $\Delta J^2$. The objective is to get new insights into the empirical behavior of volatility as it relates to signed jumps in terms of forecasting performance. Figure (8) presents the percentage signed jump for the 3 indexes considered:

$$\%\Delta J^2 = \frac{(RS^+ - RS^-)}{RV_i}$$ (1.26)

The proposed family of models can be parsimoniously estimated by OLS. Tables 11-19 show the adjusted $R^2$ as well as the estimated parameters of each model with their respective standard errors. For the 3 indexes the Time Varying versions of the models perform better. If we compare the models among them, for the case of the EUR Index and HK Index, the SJ-TVHAR model has better fit results. For the case of the USA-index, the SSJ-TVHAR has larger adjusted $R^2$. In terms of the estimation results, I have the following findings: a) SV-TVHAR. The positive semi variance has more impact in future volatility than the negative realized semi variance. We see that the decomposition of the Realized Variance adds new information since $\beta^+$ is not equal to $\beta^-$. b) In the case of the SJ-TVHAR, the coefficient of the signed jump is greater and more significant than in its no TV specification. On the other hand, the continuous variation has less explicative power if we consider the TV specification. c) In the case of the SSJ-TVHAR and SSJ-HAR tables (18) and (19) show the estimation results. The continuous variation coefficient increases as we consider
the TV component. The positive signed jump coefficient also increases, except for the case of the HK index. The positive signed jump coefficient is greater relative to the negative signed jump coefficient if we consider the TV component.

1.5.2 Out-of-Sample Forecasts

This section presents the out-of-sample forecasting results for the SV-TVHAR, SJ-TVHAR and the SSJ-TVHAR models. The models are re-estimated daily on 3 different estimation windows (90 days, 360 days and 2 years). Then, a sequence of one-day-ahead forecasts are generated. In order to compare the performances I make use of the standard Mean Squared Error (MSE):

\[
MSE = E(\hat{RV} - RV)^2
\]  

(1.27)

Figure (9) shows the MSE for each model considered in the exercise. We can observe that for all windows, the SSJ-TVHAR outperforms in terms of MSE all models. To make the comparison of different models more clear, we calculate the Relative MSE Gains (%) as we did in the simulation exercise. This result is consistent with those obtained with the simulated data. We can see that all the TV specifications perform better than the standard HAR model. For the EUR-Index the relative gains w.r.t the HAR model are between 5–10%. Similar gains are obtained with respect to the HAR-Q model. If we compare it with the standard TV-HAR model without considering semi variances or signed jumps, the gains are on average 6.5%. The relative gains are superior if we compare the family to the standard HAR-J model. In the case of the USA Index, the gains of the SSJ-TVHAR model w.r.t the standard models (HAR, HAR-Q, TV-HAR) are of the order of 19.9%, 11.8% and 10.15%, respectively. The HK-Index presents the lowest gains w.r.t to the standard models, but
they are still positive. Comparing the proposed models w.r.t their no time varying versions, all of them present better forecasts as we can appreciate in figure (9). The relative performance among the proposed family ranks the SSJ-TVHAR model as the best one.

1.5.3 Error Diagnostic

Due to the nature of the HAR specification that uses lagged values of the dependent variable as predictors there is a general problem of serial autocorrelation. For every model, the Breusch-Godfrey correlation LM test was applied confirming this issue. As recommended by Corsi (2009), the Newey-West robust standard errors were computed. One aspect we can see about the residuals for all the models consider is the presence of heteroskedasticity. The Breusch-Pagan-Godfrey test was applied to the residuals for each model. In every case this aspect is present. Further research is possible considering the presence of heteroskedasticity explicitly in the family of models presented. For every model the residuals have a mean of zero or near zero as can be noted in the figures. Given the nature of the model the Gaussian assumption is no longer valid when we work with the HAR model. The only assumption we made is that the residuals are a zero mean process. Jarque-Bera test was applied to the residuals of every model and the null hypothesis of normality was rejected in every case.

1.6 Conclusions

In this paper, I focus on modeling and forecasting the RV, and unbiased and highly efficient non parametric estimator of the return volatility. The Corsi’s (2009) HAR model and its extensions has arguably become the most used empirical tool to forecast RV in recent years. Prove of that
is its growing number of citations in the literature and its extensions including jumps (HAR-J), measurement errors considerations (HAR-Q), leverage effects and long memory, which are well known stylized facts of realized measures.

This paper analyzed the performance of volatility forecasting models. To the best of my knowledge, there is no model in the empirical literature that extends the HAR model in order to study forecasting performance taking into account these facts in a time varying coefficient environment. This work aims to fill this gap by making use of the recent concepts of Realized Semivariances and Signed Jumps. The family of 3 models proposed consist of the following ingredients: First, a time varying specification for the daily lag coefficient in the HAR model. This is motivated by the fact that the difference between long term moving average of the daily RV, which is the $RV_m$ and $RV_d$, that is $(RV_d - RV_m)$ is negative with more frequency than when the difference is positive. This creates an scenario in which RV tends to quickly revert back when there is an increase in lag $RV_d$ compared with its longer-term average level, and it tends to lower revert back when lag $RV_d$ decreased compared to $RV_m$. This fact was noticed by Li (2009) and I present evidence that this is actually the case for the 3 times series considered in the empirical analysis. Second, jump variation and continuous variation are disentangled into signed measures. As a first step, I decompose the recent quadratic variation into to signed Realized Semi variances to explore the power of explanation of future volatility. As second step I extend the model using a measure of signed jumps and finally decompose it into to components related to impacts of jumps driven by positive and negative jump variation. By combining these two ingredients, I present a family of 3 time varying signed HAR models. In order to compare their forecasting performance I computed a Montecarlo Simulation of a jump diffusion process. The models outperform their no time varying counterparts and the standard HAR, HAR-Q and HAR-J standard models. Finally, we present an empirical anal-
ysis using high frequency data of three financial indexes to test the forecasting performance of the proposed family of models. The results are consistent with those obtained via the simulation. In general, the family do better than the standard models and their no time varying counterparts. The forecasting model is very simple to implement, as it is based on the parsimonious HAR framework. The gain over the standard HAR and HAR-J models is substantial, and this is specially true when we combine the time varying context with the split signed jump measure. Natural extensions to these exercise can be done for future research. In the first place, we can consider different measures of realized semi variances. One possibility could be the novel Net Realized Variance (NRV) developed by Casas, et al. (2018). Another question to be resolved is if this new specification can be improved incorporating explicitly volatility jumps into the model at the same time that jumps in returns are considered. These questions are left for further research.
1.7 Tables and figures

1.7.1 Figures

Figure 1.4: Euro-Index and Realized Variances (2014-2020)
Figure 1.5: Euro-Index ACF of Realized Variances

(a) $RV$

(b) $RV_w$

(c) $RV_{mw}$
Figure 1.6: Hong Kong-Index and Realized Variances (2014-2020)
Figure 1.7: Hong Kong-Index ACF of Realized Variances

(a) $RV$

(b) $RV_w$

(c) $RV_m$
Figure 1.8: USA-Index and Realized Variances (2014-2020)
Figure 1.9: USA-Index ACF of Realized Variances
Figure 1.10: $\Delta J^2 = (RSV^+ - RSV^-)/RV_t$
Figure 1.11: Signed Jumps $\Delta J$ for the 3 indexes

Table 1.4: Summary Statistics for log intraday Returns

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>123,082</td>
<td>-0.000081</td>
<td>0.1377</td>
<td>-9.3271</td>
<td>1064.61</td>
<td>0.00</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>123082</td>
<td>0.00005</td>
<td>0.1402</td>
<td>-5.4897</td>
<td>597.81</td>
<td>0.00</td>
</tr>
<tr>
<td>USA Index</td>
<td>123,082</td>
<td>0.000356</td>
<td>0.10176</td>
<td>-6.2128</td>
<td>1042.86</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1.5: Summary Statistics for RV

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>1558</td>
<td>0.22678</td>
<td>0.70996</td>
<td>0.0000</td>
<td>15.8867</td>
<td>0.09528</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>1558</td>
<td>1.55472</td>
<td>4.042832</td>
<td>0.0000</td>
<td>72.9217</td>
<td>0.42986</td>
</tr>
<tr>
<td>USA Index</td>
<td>1558</td>
<td>0.25427</td>
<td>0.89624</td>
<td>0.0000</td>
<td>15.61196</td>
<td>0.07569</td>
</tr>
<tr>
<td>Obs</td>
<td>Mean</td>
<td>Std</td>
<td>Min</td>
<td>Max</td>
<td>Med</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>1558</td>
<td>0.22667</td>
<td>0.62255</td>
<td>0.01101</td>
<td>10.13999</td>
<td>0.10258</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>1558</td>
<td>1.55448</td>
<td>2.54824</td>
<td>0.02462</td>
<td>32.42365</td>
<td>0.93445</td>
</tr>
<tr>
<td>USA Index</td>
<td>1558</td>
<td>0.254302</td>
<td>0.813704</td>
<td>0.007117</td>
<td>12.208001</td>
<td>0.085583</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>1558</td>
<td>0.22667</td>
<td>0.47955</td>
<td>0.02786</td>
<td>4.39983</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>1558</td>
<td>1.5528</td>
<td>1.701783</td>
<td>0.1963</td>
<td>12.7120</td>
</tr>
<tr>
<td>USA Index</td>
<td>1558</td>
<td>0.25425</td>
<td>0.660059</td>
<td>0.01867</td>
<td>6.07929</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>min</th>
<th>max</th>
<th>med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>1558</td>
<td>0.1782</td>
<td>0.63943</td>
<td>0.000163</td>
<td>14.7537</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>1558</td>
<td>1.67998</td>
<td>3.637224</td>
<td>0.00205</td>
<td>70.79934</td>
</tr>
<tr>
<td>USA Index</td>
<td>1558</td>
<td>0.21962</td>
<td>0.7693852</td>
<td>0.00005</td>
<td>13.9304</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>min</th>
<th>max</th>
<th>med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>546</td>
<td>0.253717</td>
<td>0.92935</td>
<td>0.000242</td>
<td>14.7537</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>472</td>
<td>2.75907</td>
<td>6.00256</td>
<td>0.00452</td>
<td>70.79934</td>
</tr>
<tr>
<td>USA Index</td>
<td>548</td>
<td>0.31199</td>
<td>1.0692852</td>
<td>0.000289</td>
<td>13.9303</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Mean</th>
<th>Std</th>
<th>min</th>
<th>max</th>
<th>med</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>990</td>
<td>-0.1365</td>
<td>0.3921</td>
<td>-3.89947</td>
<td>-0.000163</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>1063</td>
<td>-1.200832</td>
<td>1.5434</td>
<td>-11.820292</td>
<td>-0.002052</td>
</tr>
<tr>
<td>USA Index</td>
<td>988</td>
<td>-0.16838</td>
<td>0.52876</td>
<td>-5.3226</td>
<td>-0.00005</td>
</tr>
</tbody>
</table>
1.7.2 In-Sample Statistics

Table 1.11: (HAR) $RV_{d,t} = \beta + \beta_d RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\beta_d$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.1587</td>
<td>0.1247</td>
<td>0.2426</td>
<td>0.5756</td>
<td>0.5452</td>
<td>4.5392</td>
</tr>
<tr>
<td></td>
<td>(0.1359)</td>
<td>(0.0436)</td>
<td>(0.0909)</td>
<td>(0.0848)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>0.477350</td>
<td>0.084852</td>
<td>0.571751</td>
<td>0.032124</td>
<td>0.2223</td>
<td>6.14305</td>
</tr>
<tr>
<td></td>
<td>(0.148921)</td>
<td>(0.030630)</td>
<td>(0.05263)</td>
<td>(0.058626)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>0.589161</td>
<td>0.164680</td>
<td>0.234008</td>
<td>0.228650</td>
<td>0.113774</td>
<td>5.5257</td>
</tr>
<tr>
<td></td>
<td>(0.132556)</td>
<td>(0.029963)</td>
<td>(0.057056)</td>
<td>(0.073557)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA Index</td>
<td>0.182559</td>
<td>0.131254</td>
<td>0.564590</td>
<td>0.090313</td>
<td>0.327528</td>
<td>4.9891</td>
</tr>
<tr>
<td></td>
<td>(0.079909)</td>
<td>(0.030832)</td>
<td>(0.051530)</td>
<td>(0.050687)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.12: In Sample (HAR-Q) $RV_{d,t} = \beta + (\gamma + \beta_q RQ_{t-1}^{1/2}) RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\beta_q$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.2817</td>
<td>0.0221</td>
<td>2.1750</td>
<td>0.2488</td>
<td>0.5903</td>
<td>0.5473</td>
<td>4.5384</td>
</tr>
<tr>
<td></td>
<td>(0.1558)</td>
<td>(0.0772)</td>
<td>(1.3527)</td>
<td>(0.0909)</td>
<td>(0.0852)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>0.5355</td>
<td>0.0152</td>
<td>0.006778</td>
<td>0.564059</td>
<td>0.0524</td>
<td>0.2259</td>
<td>6.141713</td>
</tr>
<tr>
<td></td>
<td>(0.1515)</td>
<td>(0.0461)</td>
<td>(0.0033)</td>
<td>(0.0527)</td>
<td>(0.0594)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>0.767732</td>
<td>-0.1177</td>
<td>0.00117</td>
<td>0.274626</td>
<td>0.284836</td>
<td>0.1403</td>
<td>5.4959</td>
</tr>
<tr>
<td></td>
<td>(0.1330)</td>
<td>(0.0502)</td>
<td>(0.00017)</td>
<td>(0.05649)</td>
<td>(0.072897)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA Index</td>
<td>-0.03419</td>
<td>0.850047</td>
<td>-0.002351</td>
<td>0.4324</td>
<td>-0.0790</td>
<td>0.413097</td>
<td>4.8536</td>
</tr>
<tr>
<td></td>
<td>(0.0760)</td>
<td>(0.0559)</td>
<td>(0.00015)</td>
<td>(0.0489)</td>
<td>(0.0486)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The standard errors presented in this section are Newey-West Robust as done in Corsi (2009).
Table 1.13: In Sample (TV-HAR) $RV_{d,t} = \beta + (\gamma + \alpha |RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\alpha$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.2234</td>
<td>0.0486</td>
<td>0.0088</td>
<td>0.2586</td>
<td>0.5914</td>
<td>0.5478</td>
<td>4.5372</td>
</tr>
<tr>
<td></td>
<td>(0.1402)</td>
<td>(0.0600)</td>
<td>(0.0047)</td>
<td>(0.0912)</td>
<td>(0.0851)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>0.2401</td>
<td>0.4254</td>
<td>-0.0036</td>
<td>0.5459</td>
<td>-0.05490</td>
<td>0.251176</td>
<td>6.1059</td>
</tr>
<tr>
<td></td>
<td>(0.1493)</td>
<td>0.0533</td>
<td>(0.0047)</td>
<td>(0.051822)</td>
<td>(0.058620)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>0.744038</td>
<td>-0.107097</td>
<td>0.006520</td>
<td>0.281582</td>
<td>0.292797</td>
<td>0.142474</td>
<td>5.4934</td>
</tr>
<tr>
<td></td>
<td>(0.132141)</td>
<td>0.0477</td>
<td>(0.0090)</td>
<td>(0.0565)</td>
<td>(0.07289)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA Index</td>
<td>0.0246</td>
<td>0.6775</td>
<td>-0.009637</td>
<td>0.489700</td>
<td>0.385417</td>
<td>0.385417</td>
<td>4.9023</td>
</tr>
<tr>
<td></td>
<td>(0.0776)</td>
<td>0.0546</td>
<td>(0.0081)</td>
<td>(0.0497)</td>
<td>(0.0504)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.14: In Sample (SV-HAR) $RV_{d,t} = \beta + \beta^+ R^+_{t-1} + \beta^- R^-_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\beta^+$</th>
<th>$\beta^-$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.1345</td>
<td>-0.1062</td>
<td>0.3083</td>
<td>0.3298</td>
<td>0.5224</td>
<td>0.5641</td>
<td>4.50004</td>
</tr>
<tr>
<td></td>
<td>(0.1333)</td>
<td>(0.0601)</td>
<td>(0.0574)</td>
<td>(0.0905)</td>
<td>(0.0837)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>0.401594</td>
<td>0.624112</td>
<td>0.030352</td>
<td>0.490036</td>
<td>-0.041967</td>
<td>0.2609</td>
<td>6.0955</td>
</tr>
<tr>
<td></td>
<td>(0.145632)</td>
<td>0.068468</td>
<td>(0.030541)</td>
<td>(0.052284)</td>
<td>(0.057853)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>0.597672</td>
<td>0.050538</td>
<td>0.189699</td>
<td>0.2529</td>
<td>0.24168</td>
<td>0.1171</td>
<td>5.5225</td>
</tr>
<tr>
<td></td>
<td>(0.1323)</td>
<td>(0.0527)</td>
<td>(0.01313)</td>
<td>(0.0573)</td>
<td>(0.0735)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA Index</td>
<td>0.1784</td>
<td>0.452483</td>
<td>0.030941</td>
<td>0.568912</td>
<td>-0.005006</td>
<td>0.3496</td>
<td>4.9563</td>
</tr>
<tr>
<td></td>
<td>(0.0785)</td>
<td>(0.0534)</td>
<td>(0.0332)</td>
<td>(0.0506)</td>
<td>(0.0515)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.15: In Sample (SV-TVHAR) $RV_{d,t} = \beta + (\alpha |RV_{d,t-1} - RV_{m,t-1}||RV_{d,t-1} + \beta^+ R^+_{t-1} + \beta^- R^-_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>$\beta^+$</th>
<th>$\beta^-$</th>
<th>$\alpha$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation</td>
<td>0.1813</td>
<td>-0.1549</td>
<td>0.2494</td>
<td>0.0063</td>
<td>0.3391</td>
<td>0.5350</td>
<td>0.5646</td>
<td>4.50075</td>
</tr>
<tr>
<td></td>
<td>(0.1378)</td>
<td>(0.0703)</td>
<td>(0.0700)</td>
<td>(0.0047)</td>
<td>(0.0908)</td>
<td>(0.0841)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EUR Index</td>
<td>0.2938</td>
<td>0.6635</td>
<td>0.2288</td>
<td>-0.0019</td>
<td>0.4980</td>
<td>-0.0691</td>
<td>0.2643</td>
<td>6.0888</td>
</tr>
<tr>
<td></td>
<td>(0.1483)</td>
<td>(0.0691)</td>
<td>(0.0644)</td>
<td>(0.0005)</td>
<td>(0.0521)</td>
<td>(0.0581)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>0.7449</td>
<td>-0.1021</td>
<td>-0.1112</td>
<td>0.0065</td>
<td>0.2807</td>
<td>0.2925</td>
<td>0.1419</td>
<td>5.4957</td>
</tr>
<tr>
<td></td>
<td>(0.1323)</td>
<td>(0.0567)</td>
<td>(0.0544)</td>
<td>(0.009)</td>
<td>(0.0567)</td>
<td>(0.0729)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA Index</td>
<td>0.0328</td>
<td>0.6477</td>
<td>0.6310</td>
<td>-0.0090</td>
<td>0.4945</td>
<td>-0.0788</td>
<td>0.3837</td>
<td>4.9030</td>
</tr>
<tr>
<td></td>
<td>(0.0780)</td>
<td>(0.0584)</td>
<td>(0.0724)</td>
<td>(0.0009)</td>
<td>(0.0499)</td>
<td>(0.0507)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.16: In Sample.(SJTV-HAR)$RV_{d,t} = \mu + \alpha|RV_{d,t-1} - RV_{m,t-1}| + \phi^J RV_{d,t-1} + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th>Simulation</th>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\phi^J$</th>
<th>$\phi^C$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>0.2649</td>
<td>0.1804</td>
<td>4.1507</td>
<td>-0.0023</td>
<td>0.0234</td>
<td>-0.0474</td>
<td>0.3370</td>
<td>5.9901</td>
</tr>
<tr>
<td>HK Index</td>
<td>0.6999</td>
<td>0.0040</td>
<td>-0.1859</td>
<td>0.0050</td>
<td>0.2411</td>
<td>0.2794</td>
<td>0.1421</td>
<td>5.4977</td>
</tr>
<tr>
<td>USA Index</td>
<td>-0.0086</td>
<td>0.2598</td>
<td>3.1472</td>
<td>-0.0065</td>
<td>-0.2855</td>
<td>0.1155</td>
<td>0.5015</td>
<td>4.6942</td>
</tr>
</tbody>
</table>

Table 1.17: In Sample.(SJ-HAR)$RV_{d,t} = \mu + \phi^J \Delta J_{t-1} + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th>Simulation</th>
<th>$\mu$</th>
<th>$\phi^J$</th>
<th>$\phi^C$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>0.3801</td>
<td>0.3193</td>
<td>3.4882</td>
<td>0.0404</td>
<td>-0.0231</td>
<td>0.3243</td>
<td>6.0057</td>
</tr>
<tr>
<td>HK Index</td>
<td>0.6159</td>
<td>-0.1208</td>
<td>-0.3953</td>
<td>0.3788</td>
<td>0.2415</td>
<td>0.1132</td>
<td>5.5295</td>
</tr>
<tr>
<td>USA Index</td>
<td>0.1034</td>
<td>0.4116</td>
<td>2.4059</td>
<td>-0.1861</td>
<td>0.1140</td>
<td>0.4628</td>
<td>4.7676</td>
</tr>
</tbody>
</table>

Table 1.18: In Sample.(SSJ-HAR)$RV_{d,t} = \mu + \phi^J \Delta J_{t-1}^2 + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th>Simulation</th>
<th>$\mu$</th>
<th>$\phi^J^+$</th>
<th>$\phi^J^-$</th>
<th>$\phi^C$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>0.3821</td>
<td>0.2997</td>
<td>0.3285</td>
<td>3.5303</td>
<td>0.0406</td>
<td>-0.0222</td>
<td>0.3243</td>
<td>6.0070</td>
</tr>
<tr>
<td>HK Index</td>
<td>0.5945</td>
<td>0.0420</td>
<td>-0.1956</td>
<td>-0.4540</td>
<td>0.2851</td>
<td>0.2403</td>
<td>0.1200</td>
<td>5.5231</td>
</tr>
<tr>
<td>USA Index</td>
<td>0.0985</td>
<td>0.2633</td>
<td>0.5306</td>
<td>2.6386</td>
<td>0.2048</td>
<td>0.1371</td>
<td>0.4678</td>
<td>4.7596</td>
</tr>
</tbody>
</table>

Table 1.19: In Sample.(SSJ-TVHAR)$RV_{d,t} = \mu + \alpha|RV_{d,t-1} - RV_{m,t-1}| + \phi^J \Delta J_{t-1}^2 + \phi^C BV_{t-1} + \beta_w RV_{w,t-1} + \beta_m RV_{m,t-1} + u_t$

<table>
<thead>
<tr>
<th>Simulation</th>
<th>$\mu$</th>
<th>$\phi^J^+$</th>
<th>$\phi^J^-$</th>
<th>$\phi^C$</th>
<th>$\beta_w$</th>
<th>$\beta_m$</th>
<th>$R^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>0.2058</td>
<td>0.3400</td>
<td>0.0372</td>
<td>3.9319</td>
<td>-0.0032</td>
<td>0.0157</td>
<td>-0.0649</td>
<td>0.3390</td>
</tr>
<tr>
<td>HK Index</td>
<td>0.7491</td>
<td>-0.1115</td>
<td>0.1189</td>
<td>0.0581</td>
<td>0.0067</td>
<td>0.2853</td>
<td>0.2934</td>
<td>0.1452</td>
</tr>
<tr>
<td>USA Index</td>
<td>-0.0929</td>
<td>0.5917</td>
<td>-0.2502</td>
<td>3.0746</td>
<td>-0.0122</td>
<td>-0.3151</td>
<td>0.0441</td>
<td>0.5202</td>
</tr>
</tbody>
</table>
### 1.7.3 Error Diagnostic

Table 1.20: Errors Diagnostic (SV-TVHAR) \( RV_{d,t} = \beta + (\alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \beta^{(+)}RS^+_{t-1} + \beta^{(-)}RS^-_{t-1} + \beta_wRV_{w,t-1} + \beta_mRV_{m,t-1} + u_t \)

<table>
<thead>
<tr>
<th></th>
<th>BPG</th>
<th>LM Test corr</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>119.80</td>
<td>43.07811</td>
<td>2651754</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>3.55</td>
<td>26.546</td>
<td>184761</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>USA Index</td>
<td>140.08</td>
<td>47.9318</td>
<td>3217668</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Table 1.21: Errors Diagnostic (SJTV-HAR) \( RV_{d,t} = \mu + \alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \phi^J \Delta J_{t-1} + \phi^C BV_{t-1} + \beta_wRV_{w,t-1} + \beta_mRV_{m,t-1} + u_t \)

<table>
<thead>
<tr>
<th></th>
<th>BPG</th>
<th>LM Test corr</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>72.46</td>
<td>54.2908</td>
<td>269268</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>3.1140</td>
<td>21.9122</td>
<td>1843137</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>USA Index</td>
<td>48.28</td>
<td>59.8587</td>
<td>3929345</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

Table 1.22: Errors Diagnostic (SSJ-TVHAR) \( RV_{d,t} = \mu + \alpha|RV_{d,t-1} - RV_{m,t-1}|)RV_{d,t-1} + \phi^J \Delta J^2_{t-1} + \phi^{J-} \Delta J^{-2}_{t-1} + \phi^C BV_{t-1} + \beta_wRV_{w,t-1} + \beta_mRV_{m,t-1} + u_t \)

<table>
<thead>
<tr>
<th></th>
<th>BPG</th>
<th>LM Test corr</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR Index</td>
<td>63.08</td>
<td>128.8913</td>
<td>2562877</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Hong Kong Index</td>
<td>3.037</td>
<td>24.8604</td>
<td>1848851</td>
</tr>
<tr>
<td></td>
<td>(0.0089)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>USA Index</td>
<td>42.7411</td>
<td>47.3460</td>
<td>3285764</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis are the p-values for each of the tests statistics.
Figure 1.12: Residuals USA INDEX Models
Figure 1.13: Residuals EURO INDEX Models
Figure 1.14: Residuals HK INDEX Models

(a) SVTV-HAR
(b) SJTV-HAR
(c) SSJTV-HAR
<table>
<thead>
<tr>
<th></th>
<th>SJ-TVHAR</th>
<th>SV-TVHAR</th>
<th>SSJ-TVHAR</th>
<th>SJ-HAR</th>
<th>HARQ</th>
<th>TV-HAR</th>
<th>SJ-HAR</th>
<th>TVHAR</th>
<th>SVHAR</th>
<th>SSJ-HAR</th>
<th>TVHAR</th>
<th>HAR-J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>2 years</td>
<td>7.811236</td>
<td>7.816041</td>
<td>7.539789</td>
<td>7.136301</td>
<td>7.047292</td>
<td>7.478674</td>
<td>7.4239</td>
<td>7.135894</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
<td>1 year</td>
<td>5.947027</td>
<td>5.953816</td>
<td>5.943223</td>
<td>5.95209</td>
<td>5.951592</td>
<td>5.967757</td>
<td>5.941523</td>
<td>5.968628</td>
</tr>
</tbody>
</table>

Figure 1.15: Out-Of-Sample Results
### Table 1.23: Relative Gains MSE Euro-Index

<table>
<thead>
<tr>
<th></th>
<th>SJ-HAR</th>
<th>SV-HAR</th>
<th>SSJ-HAR</th>
<th>HAR</th>
<th>HAR-Q</th>
<th>TV-HAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJ-TVHAR</td>
<td>0.01254</td>
<td>0.0574</td>
<td>0.01251</td>
<td>0.09940</td>
<td>0.09987</td>
<td>0.0649</td>
</tr>
<tr>
<td>SV-TVHAR</td>
<td>-0.0387</td>
<td>0.0076</td>
<td>-0.039125</td>
<td>0.052315</td>
<td>0.052807</td>
<td>0.064915</td>
</tr>
<tr>
<td>SSJ-TVHAR</td>
<td>0.01667</td>
<td>0.06106</td>
<td>0.01665</td>
<td>0.10318</td>
<td>0.10365</td>
<td>0.068845</td>
</tr>
</tbody>
</table>

### Table 1.24: Relative Gains MSE HK-Index

<table>
<thead>
<tr>
<th></th>
<th>SJ-HAR</th>
<th>SV-HAR</th>
<th>SSJ-HAR</th>
<th>HAR</th>
<th>HAR-Q</th>
<th>TV-HAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJ-TVHAR</td>
<td>0.00108</td>
<td>0.00882</td>
<td>0.0044</td>
<td>0.000064905</td>
<td>-0.000227</td>
<td>-0.00194</td>
</tr>
<tr>
<td>SV-TVHAR</td>
<td>0.00333</td>
<td>0.0066</td>
<td>0.0667</td>
<td>0.0028</td>
<td>0.00203</td>
<td>0.0004</td>
</tr>
<tr>
<td>SSJ-TVHAR</td>
<td>0.00441</td>
<td>0.00735</td>
<td>0.0074</td>
<td>0.00373</td>
<td>0.003126</td>
<td>0.00134</td>
</tr>
</tbody>
</table>

### Table 1.25: Relative Gains MSE USA-Index

<table>
<thead>
<tr>
<th></th>
<th>SJ-HAR</th>
<th>SV-HAR</th>
<th>SSJ-HAR</th>
<th>HAR</th>
<th>HAR-Q</th>
<th>TV-HAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJ-TVHAR</td>
<td>-0.0003</td>
<td>0.01759</td>
<td>-0.10616</td>
<td>0.0439</td>
<td>0.03974</td>
<td>0.0629</td>
</tr>
<tr>
<td>SV-TVHAR</td>
<td>-0.0778</td>
<td>0.0350</td>
<td>-0.0866</td>
<td>0.06076</td>
<td>0.0048</td>
<td>0.0629</td>
</tr>
<tr>
<td>SSJ-TVHAR</td>
<td>0.08154</td>
<td>0.17768</td>
<td>0.07410</td>
<td>0.19967</td>
<td>0.11788</td>
<td>0.1015</td>
</tr>
</tbody>
</table>
Chapter 2

Forecasting Volatility: An empirical

R-GARCH Approach for the Mexican IPC

2.1 Introduction

It is now widely acknowledged in the literature that the use of intra-daily data (namely, in the form of realized measures) can be an element of improvement of the financial volatility models (Hansen and Lund, 2011).

Realized Volatility (RV) based modeling and forecasting methods have become very popular to model using High Frequency Data. Within the approach that makes use of volatility models where the conditional variance is driven by a realized measure, the Realized GARCH model proposed by Hansen, et, al. (2012) where a measurement equation is included in the model related the realized variance to the latent conditional volatility, has become very popular in the empirical literature.

The main reason is that the R-GARCH model generally outperforms the other GARCH family models, namely the GARCH-X and HEAVY models introduced by Engle (2012) and Engle and

The first complication that arise with these approaches is the long persistence (high autocorrelations) that appears in many realized measures. Long memory or persistence is one of the most studied time series features (or stylized facts) of financial returns (Ding et al. 1993; Bollerslev and Ole Mikkelsen 1996; Breidt et al. 1998; Yalama and elik 2013). It refers to the fact that the autocorrelation of the squared or absolute returns of financial assets, as a proxy for underlying volatilities, decay at a slow rate. Different models that account for this include Baillie et al. (1996) who proposed the fractionally integrated GARCH model (FIGARCH) to account for long memory volatility and Andersen et al. (2003) who used this framework to model Realized Variance (RV). Another contribution is the Markov Switching in Maheu and McCurdy (2002). We can also consider within this literature the Unobserved Component models due to Barndor-Nielsen and Shephard (2002a) and Koopman, Jungbacker, and Hol (2005) and the Mixed DataSampling (MIDAS) approach of Ghysels et al. (2006). Another model that uses realized measures directly is Corisi’s (2009) Heterogenous Autoregressive (HAR) model that has become one of the workhorses tools for modeling RV.

More recently Huang, et, al. (2016) introduced the Corsi’s (2009) autoregressive specification to the volatility equation to account for long memory. The author documents that the standard R-GARCH is unable to capture all the long memory effects presented in several stock returns series. Empirically he is able to show that this model has superior performance.

In the same spirit, Gerlach (2016) developed a family of flexible R-GARCH models accounting for measurement error of the realized variance making use of the Realized Quarticity.
The objective of this chapter is to forecast the volatility of the IPC Mexican Index by proposing a flexibilization of the Realized GARCH model. The contribution of this paper is related to the work of Gerlach and Huang but makes even more flexible the R-GARCH model considering 3 different directions. First, I adopt an autoregressive specification of the volatility dynamics equation in the spirit of Huang, et al. (2016). Second, I allow for time varying coefficient in the volatility dynamics. With these I face the second complication we face when we want to incorporate a HAR structure to the volatility dynamics. That is, the asymmetry that exists in the relation between the realized variance and its monthly moving average. Finally, I extend the model by incorporating a jump-adjusted realized measure into the measurement equation. This allows to correct the additional bias generated by the occurrence of jumps. At the end of the paper an empirical analysis is applied to the Mexican IPC index (The Quotes and Prices Index). To the best of my knowledge this is the first work that consider this particular series in a high frequency context for forecasting. Using standard loss functions, we show that our new extension outperforms the classic Realized GARCH (1,1) model in volatility forecasting, both in-sample and out-of-sample. In section 2, I present the econometric methodology as well as the proposed new specification. In section 3, I present the data used for the empirical analysis. In section 4, I perform the in-sample and out-of-sample analysis. Finally I present the results and conclusions.
2.2 Econometric Methodology

2.2.1 Realized GARCH

The Realized GARCH (RGARCH) model introduced by Hansen et al. (2012), proposes an extension to the class of GARCH models by incorporating an explicit equation for the realized variance measure, a more efficient proxy of the volatility dynamics. With this specification the realized measure is an explanatory variable. A second extension this model considers is the addition of a measurement equation that captures the relationship between the realized measure and the latent conditional variance.

Let \( \{r_t\} \) denote a series of returns of an asset (sampled daily, for example). \( RV \) is the usual log Realized Variance, but it can be any other realized measure. Let \( \mathcal{F}_{t-1} \) denote the information set generated by all current and past information. The conditional variance is denoted as \( \sigma_t^2 = \mathbb{E}[r_t|\mathcal{F}_{t-1}] \). Define \( \log(\sigma_t^2) = h_t \) and \( \log(RV_t) = x_t \) Now, consider the standard Realized GARCH Model:

\[
\begin{align*}
    r_t &= \mu + \sqrt{h_t}z_t, \quad (2.1) \\
    h_t &= \omega + \beta h_{t-1} + \gamma x_{t-1}, \quad (2.2) \\
    x_t &= \xi + \phi h_t + \tau_1 z_t + \tau_2 (z_t^2 - 1) + u_t, \quad (2.3)
\end{align*}
\]

where \( w_t = \tau(z_t) + u_t \), with \( \tau(z) = \tau_1 z + \tau_2 (z^2 - 1) \)

The innovations \( z_t \) and \( u_t \) are identically and mutually independent with \( z_t \sim (0, 1) \) and \( u_t \sim (0, \sigma_u^2) \). The function \( \tau(z_t) \) is a measurement equation accounting for leverage effects. \( z_t \) is a standard normal random variable and \( u_t \) is a normal random variable with constant variance \( \sigma_u^2 \).
2.2.2 The Model

The forecasting methodology I propose is to adopt a time varying autoregressive specification of the GARCH equation:

\[ h_t = \omega + \beta h_{t-1} + (\gamma + \alpha |x_{d,t-1} - x_{m,t-1}|)x_{d,t-1} + \beta_w x_{w,t-1} + \beta_m x_{m,t-1}, \quad (2.4) \]

where

\[ x_{w,t} = \frac{1}{5} \sum_{i=2}^{5} x_{t-i}, \quad (2.5) \]
\[ x_{m,t} = \frac{1}{17} \sum_{i=6}^{22} x_{t-i}, \quad (2.6) \]

and the model is completed with the usual equation for the return

\[ r_t = \sqrt{h_t} z_t, \quad (2.7) \]

and with the measurement equation:

\[ x_t = \xi + \phi(\omega + \beta h_{t-1} + \gamma_d x_{t-1} + \alpha |x_{d,t-1} - x_{m,t-1}|x_{d,t-1} + \beta_w x_{w,t-1} + \beta_m x_{m,t-1}) + \epsilon_t, \quad (2.8) \]

where \( \epsilon_t = \tau(z_t) + u_t \). Then,

\[ x_t = \xi + \phi \omega + \phi \beta \left( \frac{1}{\phi} x_{t-1} - \xi - \epsilon_{t-1} \right) + \phi \alpha |x_{d,t-1} - x_{m,t-1}|x_{d,t-1} + \phi \gamma_d x_{t-1} + \phi \beta_w x_{w,t-1} + \phi \beta_m x_{m,t-1} + \epsilon_t, \quad (2.9) \]
\[ x_t = \mu_x + \beta x_{t-1} + \phi (\alpha |x_{d,t-1} - x_{m,t-1}| x_{d,t-1} + \phi (\gamma x_{t-1} + \beta w x_{w,t-1} + \beta_m x_{m,t-1}) - \beta \epsilon_{t-1} + \epsilon_t \]  

(2.10)

where \( \mu_x = \phi \omega + \xi (1 - \beta) \).

Let's look at the dynamics of the volatility equation by substituting (2.4) in the new volatility expression:

\[ h_t = \omega + \beta h_{t-1} + \gamma x_{t-1} + (\alpha |x_{d,t-1} - x_{m,t-1}| x_{d,t-1}) + \frac{\beta_w}{4} \sum_{i=2}^{5} x_{t-i} + \frac{\beta_m}{17} \sum_{i=6}^{22} x_{t-i}, \]  

(2.11)

\[ h_t = \omega + \beta h_{t-1} + \gamma (\xi + \phi h_{t-1} + \epsilon_{t-1}) + (\alpha |x_{d,t-1} - x_{m,t-1}| x_{d,t-1}) + \frac{\beta_w}{4} \sum_{i=2}^{5} (\xi + \phi h_{t-i} + \epsilon_{t-i}) + \frac{\beta_m}{17} \sum_{i=6}^{22} (\xi + \phi h_{t-i} + \epsilon_{t-i}), \]  

(2.12)

\[ h_t = \omega + \gamma \xi + h_{t-1} (\beta + \gamma \phi) + (\alpha |x_{d,t-1} - x_{m,t-1}| x_{d,t-1}) + \beta_w \xi + \beta_m \xi + \frac{\phi \beta_w}{4} \sum_{i=2}^{5} h_{t-i} + \frac{\phi \beta_m}{17} \sum_{i=6}^{22} h_{t-i} + \frac{\beta_w}{4} \sum_{i=2}^{5} \epsilon_{t-i} + \frac{\beta_m}{17} \sum_{i=6}^{22} \epsilon_{t-i} + \gamma \epsilon_{t-1}, \]  

(2.13)

which is a corrected ARMA specification. Note that the extension that has been presented so far does not consider the presence of jumps. I consider another extension in the spirit of Gerlach, et.al. (2018). The objective is to include this other source of bias in the model influenced by jumps. What I consider is a modification of the original measurement equation (3) by adding the log-ratio between the Realized Variance and a jump-robust realized measure \( x_t^J \) as an explanatory variable. For that purpose I use the Bivariate Variation (BV).

\[ BV_t = \frac{\pi}{2} \sum_{i=2}^{m} |r_{t,i}||r_{t,i-1}| \]  

(2.14)
Let $S_t = \frac{x_t^2}{x_t^{1.5}}$. This ratio converges to $\frac{QV}{IV}$. Values of $S_t > 1$ is evidence of the occurrence of jumps at time $t$. $S_t$ can be interpreted as a bias correction term that works in the following direction via the measurement equation:

$$
\log(x_t) = \xi + \delta \log(h_t) + \eta \log(S_t) + \tau(z_t) + u_t^*, 
$$  \hspace{1cm} (2.15)

that is:

$$
\log(x_t^*) = \xi + \delta \log(h_t) + \tau(z_t) + u_t^*,
$$ \hspace{1cm} (2.16)

where $\log(x_t^*) = \log(x_t/S_t^n)$. With this new measurement equation the dynamics of the volatility can be seen in the following equation:

$$
h_t = \omega + \gamma \xi + h_{t-1}(\beta + \gamma \phi) + (\alpha|x_{d,t-1} - x_{m,t-1}|x_{d,t-1}) + \beta_w \xi + \beta_m \xi + \frac{\phi \beta_w}{4}\sum_{i=2}^{5} h_{t-i} + \frac{\phi \beta_m}{17}\sum_{i=6}^{22} h_{t-i} + \frac{\beta_w}{4}\sum_{i=2}^{5} \epsilon_{t-i}^* + \frac{\beta_m}{17}\sum_{i=6}^{22} \epsilon_{t-i}^* + \gamma \epsilon_{t-1}^*, \hspace{1cm} (2.17)
$$

where $\epsilon_t^* = \tau(z_t) + u_t^*$. In order to get the value of $\eta$, we look at the following relation:

$$
\log(x_t^*) = \log(x_t) - \eta \log(S_t), \hspace{1cm} (2.18)
$$

$$
= \log(x_t) - \eta[\log(x_t) - \log(x_t^*)], \hspace{1cm} (2.19)
$$

$$
= (1 - \eta) \log(x_t) + \eta \log(x_t^*). \hspace{1cm} (2.20)
$$

As $\eta$ lies between 0 and 1, this parameter is a weight assigned to the log jump-robust realized measure. In the following section, I perform an empirical analysis using these proposed models for
the mexican IPC case.

2.2.3 Data

I use data of daily realized variance of the major financial index in Mexico called the IPC (Quotes and Prices index). The realized measure is constructed from intraday data sampled every 5 minutes. The sample spans the period from January 3, 2000 to December 5, 2017, implying a total of 4486 observations. The data is obtained from the Realized Library Dataset of the Oxford-Man Institute of Quantitative Finance.

Daily returns are constructed for the same time period using the average price using the open and close data. IPC is the main stock exchange of the mexican financial market.

It is constructed with the principal 35 firms listed in the market. Figure 1 shows the data corresponding to the Mexican IPC Realized Variance and the corresponding Autocorrelation Function for the log-transformed data. We can note the classic high autocorrelation even for long lags.

Table 1 shows the descriptive statistics of the data corresponding to the daily returns and log($RV$).

All the returns show fat tails with large kurtosis. This characteristic of the returns are not considered in the present analysis.
Figure 2.1: Mexican IPC- RV and ACF (2000-2017)
The choice to use RV sampled at 5 minutes is related to the general consensus in the literature that such frequency eliminates related microstructure noise problems in the data. Considering log$(RV)$ I estimate the fractional integrated parameter $d$ for the entire sample. The Local Whittle estimator is used as proposed by Robinson (1995). The results for the Mex-IPC log Realized Variance is $\hat{d} = 0.62 > 1/2$ with a standard deviation of 0.05. This indicates the long memory property of the series. Table 1 presents the summary statistics of the data and the daily returns series.

<table>
<thead>
<tr>
<th>Table 2.1: Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Returns log$(\tilde{RV})$</td>
</tr>
<tr>
<td>Obs  4486  04486</td>
</tr>
<tr>
<td>Mean 0.0003677 -9.925</td>
</tr>
<tr>
<td>Std  0.012787  0.94945</td>
</tr>
<tr>
<td>Min -0.08261 -12.601</td>
</tr>
<tr>
<td>Max  0.0995268 -5.257</td>
</tr>
<tr>
<td>Skew -0.0003915 0.6711722</td>
</tr>
<tr>
<td>Kurt 5.2843 0.745661</td>
</tr>
</tbody>
</table>

In figure 2 we can see the graph for the adjusted variable $S_t = \frac{x_t}{x^*_t}$ and the log$(\frac{x_{t+1}}{S_t}) = \log(x^*_t)$. For the case of the Mexican IPC, according to the empirical analysis, $\eta = 0.00006176$. Also, we can observe the classic problem of asymmetry between $RV_{d,t}$ and $RV_{m,t}$. 

53
Figure 2.2: Mexican IPC- $S_t$ and Asymmetry

(a) $S_t = \frac{x_t}{x_t'}$

(b) Blue = $RV_{d,t}$ Red = $RV_{m,t}$
55% of the time $RV_{d,t} < RV_{m,t}$
2.2.4 Estimation

Parameters can be estimated by standard QMLE techniques given that the conditional density \( f(r_t, x_t|\mathcal{F}_{t-1}) \) can be decomposed into:

\[
f(r_t, x_t|\mathcal{F}_{t-1}) = f(r_t|\mathcal{F}_{t-1})f(x_t|r_t; \mathcal{F}_{t-1}). \tag{2.21}
\]

Following Hansen et al. (2012), the quasi log-likelihood function can be written as:

\[
\mathcal{L}(r, x; \tilde{\theta}) = \sum_{t=1}^{T} \log f(r_t, x_t|\mathcal{F}_{t-1}), \tag{2.22}
\]

where \( \tilde{\theta} = (\bar{\theta}_h, \bar{\theta}_x, \bar{\theta}_\sigma)' \), with \( \bar{\theta}_h, \bar{\theta}_x, \bar{\theta}_\sigma \) respectively being the vectors of parameters in the volatility equation, in the measurement equation and in the noise variance specification. Assuming a Gaussian specification of \( z_t \) and \( u_t \), the quasi log-likelihood function is:

\[
\mathcal{L}(r, x; \theta) = -\frac{1}{2} (\log(2\pi) \sum_{t=1}^{T} \log(h_t) + \frac{r_t^2}{h_t} + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}). \tag{2.23}
\]

2.2.5 In-Sample Results

I present the results for the Mexican IPC index using the daily returns and log \( RV \) for the 3 models considered covering the entire sample period:
• **Realized GARCH**

\[ r_t = \sqrt{h_t} z_t, \quad (2.24) \]

\[ h_t = -0.426478 + 0.7175 h_{t-1} + 0.212613 x_{t-1}, \quad (2.25) \]

\[ x_t = 0.765813 + 1.190546 h_t - 0.047546 z_t + 0.127731 (z_t^2 - 1) + u_t, \quad (2.26) \]

BIC: -4.3584

\[ \mathcal{L}(r, x) : 9809.633 \]

• **Realized HAR Garch**

\[ r_t = \sqrt{h_t} z_t, \quad (2.27) \]

\[ h_t = -0.449509 + 0.691913 h_{t-1} + 0.223096 x_{t-1} + \]

\[ -0.008747 \left( \frac{1}{4} \sum_{i=2}^{5} x_{t-i} \right) + 0.020130 \left( \frac{1}{17} \sum_{i=6}^{22} x_{t-i} \right), \quad (2.28) \]

\[ x_t = 0.703426 + 1.183479 h_t + 0.047615 z_t + 0.128166 (z_t^2 - 1) + u_t, \quad (2.29) \]

BIC: -4.3616

\[ \mathcal{L}(r, x) : 9825.146 \]
**Realized TV-HAR Garch**

\[ r_t = \sqrt{h_t} z_t, \]  
\[ (2.30) \]

\[ h_t = -0.459009 + 0.658839 h_{t-1} + 0.229919 x_{t-1} + \]
\[ + 0.001725 (|x_{d,t-1} - x_{m,t-1}|) x_{d,t-1} + \]
\[ -0.000555 \left( \frac{1}{4} \sum_{i=2}^{5} x_{t-i} \right) + 0.032194 \left( \frac{1}{17} \sum_{i=6}^{22} x_{t-i} \right), \]
\[ x_t = 0.700053 + 1.183205 h_t + -0.047789 z_t + 0.128677 (z_t^2 - 1) + u_t, \]  
\[ (2.31) \]

**BIC:** -4.3608

\[ L(r, x): 9827.609 \]

**RJ-Adjusted TV-HAR-GARCH**

\[ r_t = \sqrt{h_t} z_t, \]  
\[ (2.33) \]

\[ h_t = -0.462624 + 0.692781 h_{t-1} + 0.220139 x_{t-1}^* + \]
\[ + 0.003534 (|x_{d,t-1}^* - x_{m,t-1}^*|) x_{d,t-1}^* + \]
\[ -0.008615 \left( \frac{1}{4} \sum_{i=2}^{5} x_{t-i}^* \right) + 0.021869 \left( \frac{1}{17} \sum_{i=6}^{22} x_{t-i}^* \right), \]
\[ x_t^* = 0.680733 + 1.180668 h_t + -0.047536 z_t + 0.127870 (z_t^2 - 1) + u_t, \]  
\[ (2.34) \]

**BIC:** -4.3624

\[ L(r, x): 9831.188 \]

---

57
The numbers in parenthesis are the robust standard errors for the estimates. The first finding is that the parameter $\beta$ is lower than the R-GARCH and the R-HAR-GARCH models when the persistence parameter increases. The significance of the parameters in these nested models are tested by the Log Likelihood test, that is:

$$LR = 2(\log L_u - \log L_r),$$

where $u$ and $r$ stand for the unrestricted and restricted models. Since the adjusted for jumps model is not nested with the other 3, the LR test is not computed. Note that the Log-Likelihood of the jumps-adjusted model is the higher of all four.
2.2.6 Out-of-Sample Results

This section presents the out-of-sample performance (predictive ability) of the models. I use a standard rolling window scheme, with a window size of 2693 days. Hence, the out-of-sample period includes 1793 observations. I evaluate the forecast performance of the models using the following 2 loss functions:

\[ QLIKE = \frac{1}{T} \sum_{t=1}^{T} \log(\hat{h}_t) + \frac{x_t}{\hat{h}_t^2}, \]  

(2.37)

\[ MSE = \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{h}_t)^2, \]  

(2.38)

where \( RV \) is used as a volatility proxy. The results are presented in Table 4. The results show that the adjusted for jumps model dominates the other 3 for the 2 class of loss functions considered. The gains are particularly larger between the HAR-GARCH class and the original R-GARCH model.

<table>
<thead>
<tr>
<th>Model</th>
<th>QLIKE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGARCH</td>
<td>-1.700768</td>
<td>0.0004190</td>
</tr>
<tr>
<td>R-HAR GARCH</td>
<td>-1.804008</td>
<td>0.00026692</td>
</tr>
<tr>
<td>TV-HAR-RGARCH</td>
<td>-1.8015</td>
<td>0.0002638.</td>
</tr>
<tr>
<td>RJ-Adjusted HAR GARCH</td>
<td>-1.8452</td>
<td>0.0002155</td>
</tr>
</tbody>
</table>

2.2.7 Conclusions

This paper proposes a generalization of the Realized GARCH model of Hansen, et.al. (2012) in order to estimate and forecast the Mexican IPC (Quotes and trade index) using High Frequency Data. The more flexible specification allows at the same time consider 3 characteristic of the data: i) the high persistence or long memory presented in the IPC. By adopting a HAR specification in the spirit of Zhuo, et.al. (2015) of the volatility dynamics equation we incorporate this feature of
the data. ii) The asymmetry we observe between the Realized Variance and its Moving Average. When we work with the Corsi’s (2009) HAR model, the parameter of the first lag vary over time. This feature is considered by adopting a time varying coefficient environment in the AR extension of the volatility equation. iii) The occurrence of jumps. By including a jump-adjusted realized measure in the measurement equation in the spirit of Gerlach, et.al. (2018) we correct for the bias generated by jumps. The empirical analysis is focused on the estimation and forecasting of the Mexican IPC index. This series has all the classical features of the Realized Variance. The results show that our modelling exercise outperforms the standard Realized GARCH both in fitting and forecasting volatility. Extensions to this model can be considered in several directions: i) Considering heteroskedasticity of the error term in the measurement equation, ii) Consider measurement errors, iii) extend the analysis to another financial variables (exchange rates, for example).
Chapter 3

Income Inequality and free trade: An
applied causality analysis for Mexico

3.1 Introduction

Causality is one of the main concerns when we analyze relationships between economic variables. The develop of causality analysis in economic variables has a direct implication in public policy and forecasting. Perhaps the most influential definition of causality in economics comes from the work of Granger (1969) who defined causality of a variable $x_t$ to $y_t$ if $x_t$ contains past information that can be used to predict the contemporary $y$. In this paper, the objective is to establish if there are causality relationships in the sense of Granger between income inequality and trade in Mexico.

In this sense and as a motivation it is important to notice that two of the most well known phenomena in the world economy in recent years have been the interconnection among countries and a significant rise in inequality. What is the relationship between globalization and income inequality in Mexico? In the last decade there has been an intense debate about the impact of the
Mexican trade liberalization occurred in the early eighties on economic and social performance. Figure 1 shows the evolution of the income inequality and the trade integration in Mexico since 1963. It can be observed that income inequality (measured by the GINI Coefficient of the Total Disposable Income) had a decreasing trend between 1963 and 1983 and the presented an increasing trend between 1983 and 1998 and declined between 1998 and 2014 to finally reach the same levels presented in the early eighties. At the same time, trade integration has suffered a clear increment over the years. This paper is concerned with the following question: is there a causal relationship between inequality and trade volume in Mexico? Mexico began the economic liberalization process in the late 80’s with the country’s entry into GATT and consolidated it with the firm of NAFTA (joint with USA and Canada) in 1994. Nowadays Mexico has signed over 40 free trade agreements with different countries which is reflected in the fact that the trade by GDP ratio is almost 0.55 (see figure 1). At first glance it would be fair to presume that trade liberalization is not enough to reduce income inequality over time (or at least it has not given the expected results posed prior the trade liberalization). In order to give some perspective of the economic context in Mexico, let me present some relevant data. Mexico is nowadays the 14th economy in the world in terms of GDP (1143.79 billion US dollars) and with a GDP per capita of 10,307 USD dollars­just about the world average of 10,057 US dollars (World Bank)­. That is, the GDP per capita is more than 5 times that of 1983. However, since the implementation of the neoliberal reforms along with the strong increment in exports of manufactures, Mexico’s economic growth - an annual average of 2.3% - has been very low relative to the growth rates seen in the 70’s and early 80’s (Esquivel, G. 2015). In Mexico more than 55 million people live in poverty (that is almost the 50% of the population), a proportion similar to the one seen 30 years ago (CONEVAL, Mexico). On the other hand, Mexico is ranked 87 among 113 countries in terms of income inequality (being the first
one the least unequal), that is, Mexico falls within the 25% of countries with the highest levels of inequality in the world. Looking at the data, it seems that Mexico faces a context characterized by stable economic growth (between 2003 and 2014 Mexican economy grew 2.6% on average), a monotone increment in trade integration, persistent levels of poverty and a wide income inequality relative to the average (measured by the standard GINI Index). There is a vast literature that tried to explain the "hump" form of the GINI index in Mexico the years that followed trade liberalization. That may be the consequence of various factors, including changes in the skill premium, social policies (basically government transfers) and hours worked (Krozer, A and Moreno, J. 2014). In spite of the latter, income inequality in Mexico is still very high. In fact, "despite the significant increase in social spending as a share of GDP, low inflation, the implementation of NAFTA and of a series of radical reforms Mexico has been unable to significantly reduce its high concentration of income" (Krozer, A and Moreno, J. 2014). There is no doubt that income inequality and poverty
are structural problems in developing countries that have severe implications on welfare and thus are priority issues of public policy. While this situation is by itself very challenging in terms of social welfare, the wealth inequality represents a bigger and more complicated issue in countries such as Mexico. The Global Wealth Report 2014 from Credit Suisse (2014) points out that the wealthiest 10% of people in Mexico concentrate 64.4% of the country’s total wealth while 1% of the population concentrates almost 50% of the wealth! Figure 3 (Based on Esquivel 2015) shows the income average growth rate of the top 1, 5, 10 and 50 wealthiest people in Mexico from 2005 to 2012. This figure illustrates that the dynamic of the income inequality is totally biased to the top 10% of the wealthiest people in the country, which in the long run translates to wealth inequality. This dynamic is reflected in the fact that the number of millionaires in Mexico grew by 32% between 2007 and 2012. For instance, in 1996, the wealth of the 15 individuals with fortunes over 1 billion dollars were equal to 25.6 billion dollars while that of the 16 wealthiest Mexicans in 2014 was equivalent to 142.9 billion dollars. By 2015 the fortune of the 4 main Mexicans multimillionaires is equivalent to almost 8.5% of the GDP. In this paper, we look for statistical evidence of causality and long term relationships between income inequality and trade in Mexico using the Granger causality perspective. There are many studies in the literature accounting for the correlation between these two variables, nevertheless accounting for causality or a strongest link is more difficult. In the next section, I summarize some of the work that has been made in relation with the subject of study. The topic is relevant by itself as the government of Mexico is still implementing privatizations and liberalization of strategic sectors (oil, electricity,...) and is facing an imminent renegotiation of NAFTA.
3.2 Literature Review

Many studies over the last decade have tried to explain the relationship between trade integration and income inequality in Mexico. The first study of that nature is Feenstra and Hanson (1997) who studied the effects of foreign direct investment on wage inequality in Mexico. They showed that rising wage inequality in Mexico is linked to foreign capital inflows throughout the skilled labor share of wages over the period 1975-1988. That is, FDI is positively correlated with the relative demand for skilled labor and it can account for a large portion of the increased in the skilled labor share of total wages?. This is one of the first papers that tried to find some causal relationships between economic liberalization (as a general term) and wages disparities.

Not particularly for the case of Mexico, Grossman and Helpman (1994) construct a multi-sector economy and study the case in which special interest groups lobby the own-welfare maximizing government for trade protection. This subsequently leads to income inequality. This is one of the studies that deals with the institutional framework and trade liberalization. The implications turn out to be very interesting since they have applications to study developing countries. This is
directly related to the role played by foreign investment on the power clusters formation and hence on inequality.

Rodriguez-Lopez and Esquivel (2003) applied the wage approach suggested by Leamer (1998) "to separate out the effects of trade and technology in wage inequality in Mexico". The analysis shows that technology is the principal factor that accounted for the increase in wage inequality before NAFTA. The paper suggests that without technological change, trade liberalization would have reduced wage gap in Mexico. They found that the Stolper-Samuelson result holds for the pre-NAFTA period and that the technological change has increased the returns to skills.

Faber (2014) found evidence that the "NAFTA’S relative price effect of higher quality products in mexican cities has led to a significant increase in mexican real income inequality due to differences in cost of living inflation among poor and rich households”.

Herzer, Huhne and Nunnenk (2012) perform a panel cointegration analysis to determine causality relations between foreign direct investment and income inequality in several Latin American countries. The analysis showed a positive effect on income inequality and on the other hand ”FDI contributed to widening income gaps in all individual sample countries”. This empirical analysis coincides with the before mentioned Hanson and Feenstra study.

Gobbin and Rayp (2008) apply Johansen’s cointegration methodology in order to find long term relationships between inequality and economic growth in Belgium, the US and Finland.

Other empirical studies that have tried to explain the relationship between income inequality and globalization are Rivas (2007), Sanchez-Reaza and Rodriguez-Pose (2002); Chiquiar Cikurel (2002) and Puga (1999).

Some other studies have aimed to analyze the causality between income inequality and economic growth, but the relationship between the former and trade has been less studied in the empiri-
cal literature, particularly for the case of Mexico. Example of the former studies are Chakrabarti(2000) who examined the effect of trade on intra national distribution of income and it is tested in an instrumental variable estimation of cross country regressions for 73 countries. The analysis found “a robust and statistically significant inverse relationship between trade and income inequality”. Beaton , Cebotari and Komaromi (2017) who used panel data regressions showing that “trade openness can promote economic growth without adversely affecting income inequality”. On the other hand Heshmati(2003), in a regression analysis investigates the relationship between inequality and globalization. The author showed that globalization explains between 7 and 11 percent of the variations in income inequality among the countries.

In general the literature offers a variety of results depending on the variables used to measure inequality and the methodology used. In this sees, this paper contributes with this empirical literature and tries to give an statistical explanation about the relationship between trade and income inequality in Mexico.

3.3 Methodology and data

In this paper the causality links between the trade integration and income inequality are analyzed. The Trade/GDP ratio was constructed with data from the Mexican National Institute of Geography and Statistics (INEGI). The frequency of the data is annual and the time span considered covers the years from 1963 to 2014.

For income inequality data I use first the Standardized World Income Inequality Database version 6.0 by Solt (2017). From that source I use the measure equivalent to a Total Disposable Income Gini Index (that is, it includes labor and non-labor monetary income and transfers)
In order to complete the empirical analysis, as a second step, I use the dataset computed by Esquivel, Campos and Lustig (2011). They calculate different Gini coefficients for the case of Mexico. I use these calculations to apply the cointegration and causal analysis. The first measure is the Gini coefficient for labor income. For the second measure, I use the Hourly Wage Gini coefficient which is equal to monthly labour income over weekly hours of work times 4.33. Esquivel, Campos and Lustig (2011) calculate this variable for individuals within the 18-65 years old rank with positive income and it includes labour income from wages and self-employment. In figure 3 we can see the trends followed by the time series considered in the analysis. The objective of the empirical analysis in this paper is to establish Granger causality relationships between the variables considered. In the causality analysis we take the null hypothesis that "Y does not cause X", given other variables and is proved using the Ghartey (1993) F statistic. In the first place we check the order of integration of the series by looking for the presence of unit roots. If the series are non
stationary, the second objective is to determine the existence of long term relationships (cointegration analysis) in the data. The unit root tests used in the analysis are the Augmented Dickey Fuller (ADF) and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) tests. As pointed out by Toda and Phillips (1993), the autoregressive models are not reliable when integrated series are used for estimating causal relations. In the presence of non stationarity of the data, I follow the recommendation of Naka and Tufte (1997) and an error Vector Error Correction Model (VEC) is adjusted. In the case the series are non stationary and they are not cointegrated, I estimate the causal relationships on a VAR in differences avoiding the presence of spurious results. The test of Johansen (1990) is used to determine the presence of such long term relationships. In mathematical notation the econometric technique used in this paper can be summarized in the following way. We can consider an autoregressive model of the form:

$$y_t = a + Q_1 y_{t-1} + \ldots + Q_t y_{t-p} + e_t,$$  \hspace{1cm} (3.1)

where $y_t$ is an $n$ column vector of random variables. If the variables turned out to be non stationary and cointegrated, we can formulate a Vector Error Correction model (VEC) as:

$$\Delta y_t = a + \Phi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_{t-1} \Delta y_{t-i} + e_t,$$  \hspace{1cm} (3.2)

where $\Phi = \sum_{i=1}^{p} Q_i - I$ and $\Gamma_{t} = -\sum_{j=i+1}^{p} Q_j$. The rank of $\Phi$ determines the number of linear independent cointegration vectors among the components of $y_t$. 

69
3.4 Results

In the first place, the unit root tests were applied to the series. The two unit root tests applied (Augmented Dickey Fuller and Phillips Perron) suggest that all the series present non stationarity in variance. Tables A, 1a and 1b show the statistical results. Given that the series are non stationary, we perform the Johansen’s test for cointegration. First, I consider the SWIID data set and the Trade-GDP ratio to perform the analysis.

Tables A and B show the result of the cointegration test using the Max-Eigenvalue and Trace criteria. The evidence suggests no cointegration relationships. That is, considering the entire sample (1963-2015) there are evidence to conclude that there do not exist long-term relationships between the two series. Statistical results are summarized in tables A and B. In tables A and B the null hypothesis $H_0^*$: ”There are not cointegration equations” cannot be rejected at the 5% level of significance. Given that the two variables are not cointegrated I perform the Granger causality analysis on a VAR in differences to avoid spurious conclusions on the null hypothesis. The results are shown in table C. The evidence shows that there is unidirectional Granger causality from trade-GDP ratio to the Total Disposable Income Gini Index. that is, with the past information of trade in Mexico, we are able to explain the contemporaneous Total Disposable Income inequality.

<table>
<thead>
<tr>
<th>Table A: Cointegration Test (Trace)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>----------------------------------</td>
</tr>
<tr>
<td>Trade/GDP-SWIID Gini</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table B: Cointegration Test (Max-Eigenvalue)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>----------------------------------</td>
</tr>
<tr>
<td>Trade/GDP-SWIID Gini</td>
</tr>
</tbody>
</table>
In the second place, I use Esquivel, Campos and Lustig (2011) computations. This data cover the period from 1989 to 2010. Tables 2a and 2b show the result of the cointegration test using the Max-eigenvalue and trace criteria. The evidence shows that there exists a quadratic cointegration equation between trade and labor income inequality, that is, the two variables share a common quadratic long run trend.
Since both series are non stationary and are cointegrated, I performed the Granger causality analysis on a VEC. The statistical results are displayed in table 3. The results show no evidence for granger causal relationship between trade and labor income inequality. On the other hand, there exists a long run relationship between trade and hourly wage inequality. In this case we find two cointegration equations. The results show that there in a bidirectional Granger causality between trade and the Hourly Wage Gini Index.
Finally, I perform an impulse response analysis in order to infer the impact of one variable on the other in the VAR system. In the Appendix Fig. 4 shows the effect of a Cholesky one standard deviation innovation on the Labor Income Gini Index. Given a positive shock, this variable shows negative response, slightly decrease drink the first 4 years and then shows a smooth increment. Figure 5 shows the impulse response function of the Hourly Wage Gini Index to trade-GDP ratio. For all the sample the response is negative, but it is increasing after the year 3. The path is similar to that of the labor income Gini but the response is greater. This would explain in part the relative flatness of the GINI index after the mid 80’s. Finally, in figure 6 we can see the response function of the Gini Index to trade using the Total Disposable Income data from SWIID (2016). There is negative and decreasing response during the first 2 years and then there is a positive and increasing trend and finally the function gets stable from the sixth year and beyond. In general, we can note that the income inequality tends to have a reduction in the first years after trade liberalization took place but after that short period of time, it increases smoothly again.
3.5 Conclusion

The paper has two objectives. The first one is to give an answer to the question: Is there a statistical causal relationship between trade and inequality in Mexico? The second one is to use the Granger Causality methodology and the Johansen’s cointegration model to obtain empirical results. The main findings of the paper are: a) There is a Granger Causal relationship between trade/GDP ratio and income inequality measured by the Total Disposable Income Gini Index using the SWIID data base over the sample covering the period 1963-2014. The causality analysis was performed on a VAR in differences since we did not find any long run common trend b) Using the computations of Esquivel, Campos and Lustig (2011), we conclude that there is a bidirectional causal link in the sense of granger between trade and income inequality measured by the Hourly Wage Gini Index. A quadratic cointegration equation was found. On the other hand, we did not find any causal relation between trade and income inequality measured by the Labor Income Gini Index. A quadratic cointegration equation was found. The results were not homogenous among the different measures of income inequality showing that they are very sensitive to measure mechanics. What we can conclude from the exercise is that trade has a direct and causal impact on disposable income inequality and on hourly wage inequality. In order to infer the sign of the causality, I used the impulse response functions. Why are these results important? Mexico’s development project has as one of the principals drivers (along with the trade liberalization) a rapid process of privatization. We can distinguish 3 periods that characterize this phenomenon in Mexico. The first period occurred from 1982 to 1988 just when Mexico did entry into the GATT. This period was characterized by the privatization and mergers of firms in different sectors of the economy (financial sector, fishing industry, automobile industry). The second period (1988-1994) was the most intense period of pri-
vatization. The state gave the control of important economic sectors and created important groups of power (Gonzalez, 2010).

The result is that a significant portion of the fortunes of the four richest individuals is derived from sectors that have been privatized, concessioned and/or regulated by the public sector. What are the economic forces that have driven this behavior?

How can we reconcile in a model a context in which the trade (exports and imports) increases significantly, wealth and income inequality rise and the economic and political institutions collude to concentrate the power in the economy? That would be a potential question for further investigation.
### 3.6 Statistical Results

#### 3.6.1 Unit Root Tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>t-statistic</th>
<th>p-value</th>
<th>t-statistic $(\Delta y_t)$</th>
<th>p-value $(\Delta y_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade/GDP</td>
<td>-0.972199</td>
<td>0.3432</td>
<td>-3.807695</td>
<td>0.0015</td>
</tr>
<tr>
<td>SWIID Gini</td>
<td>-2.03684</td>
<td>0.2707</td>
<td>-7.48599</td>
<td>0.0000*</td>
</tr>
<tr>
<td>LaborIncomeGini</td>
<td>-0.446377</td>
<td>0.6604</td>
<td>-3.6184</td>
<td>0.0023</td>
</tr>
<tr>
<td>HourlyWageGini</td>
<td>-2.0019</td>
<td>0.2836</td>
<td>-2.9899</td>
<td>.0573*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>t-statistic</th>
<th>p-value</th>
<th>t-statistic $(\Delta y_t)$</th>
<th>p-value $(\Delta y_t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trade/GDP</td>
<td>-0.901272</td>
<td>0.7673</td>
<td>-3.9482</td>
<td>0.0074</td>
</tr>
<tr>
<td>SWIID Gini</td>
<td>-1.782128</td>
<td>0.3850</td>
<td>-7.56723</td>
<td>0.0000*</td>
</tr>
<tr>
<td>LaborIncomeGini</td>
<td>-0.4463</td>
<td>0.8835</td>
<td>-3.0569</td>
<td>0.0466</td>
</tr>
<tr>
<td>HourlyWageGini</td>
<td>-1.73315</td>
<td>0.4011</td>
<td>-7.39582</td>
<td>0.0000*</td>
</tr>
</tbody>
</table>
3.6.2 Impulse Response Functions

Figure 3.4: Impulse Response functions Trade/GDP and Labor Income Gini

Figure 3.5: Impulse Response functions Trade/GDP and Hourly Wage Gini
Figure 3.6: Impulse Response functions Trade/GDP and Total Disposable Income Gini (SWIID)
Bibliography


[57] INEGI.

http://www.inegi.org.mx


[61] Komaromi, Beaton, Cebotari and Komaromi (2017). Revisiting the Link between Trade, Growth and Inequality: Lessons for Latin America and the Caribbean. International Monetary Fund


[80] Sánchez-Reaza and Rodríguez-Pose (2002). The Impact of Trade Liberalization on Regional Disparities in Mexico. Growth and Change, Volume 33, Issue 1


[84] Tang, L. 2015 Top Income Inequality, Aggregate Saving and the Gains from Trade.

https://ideas.repec.org/jmp/2014/pta581.pdf


http://data.worldbank.org
