THREE ESSAYS ON MACROECONOMETRICS

YUSUF VARLI
110806008

İSTANBUL BİLGİ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMİ DOKTORA PROGRAMI

DR. ORHAN ERDEM
2013

Three Essays on Macroeconometrics

Makroekonometri Üzerine Üç Çalışma

Yusuf Varlı
110806008

Tez Danışmanı: Dr. Orhan Erdem

Jüri Üyesi: Doç. Dr. Koray Akay

Jüri Üyesi: Doç. Dr. Erk Hacıhasanoğlu

Jüri Üyesi: Doç. Dr. Gökhan Özertan

Jüri Üyesi: Yrd. Doç. Dr. Serda Selin Öztürk

Tezin Onaylandığı Tarih: 06/12/2013

Toplam Sayfa Sayısı: 110

Anahtar Kelimeler (Türkçe):

Konut Arzı, Konut Talebi, Eşbütünleşme, VECM, Türkiye, Devlet Kredi Dereceleri, Sıralı Yanıt Modelleri, Pearson Korelasyon Katsayısı, Durağanlık, Durağan Olmayan Zaman Serileri

Anahtar Kelimeler (İngilizce):

Housing Supply, Housing Demand, Cointegration, VECM, Turkey, Sovereign Credit Ratings, Ordered Response Models, Pearson's Correlation Coefficient, Stationarity, Non-stationary Time Series

ABSTRACT

THREE ESSAYS ON MACROECONOMETRICS

Yusuf Varlı

Ph.D. in Department of Economics

Supervisor: Dr. Orhan Erdem

December, 2013

This dissertation consists of three essays on the area of macroeconometrics. The first essay models the demand and supply sides of the housing market separately. In the empirical analysis, the vector error correction model (VECM) is preferred to identify the impacts of fundamental macroeconomic factors on the demand and supply sides of the Turkish housing market between the period of October 2007 and June 2012. It is found that the housing market and macroeconomic fundamentals, such as the interest rate, gross domestic product (GDP) and housing prices are cointegrated. In light of the evidence on two cointegrating equations, the error correction model is estimated to examine the effect of the variables on housing demand and supply in Turkey. While the dependent variable on the demand side is real mortgage credit volume, the dependent variable explaining the supply side is the number of construction permit. The study reveals that the macroeconomic variables have different impacts on the dynamic behavior of mortgage credit and construction permit numbers. Additionally, the impulse response analysis based on structural VECM suggests that the housing market in Turkey is sensitive to shocks in the economy. This essay also presents forecast error variance decompositions (FEVD) and indicates the important role of real house prices and real GDP per capita on the housing market in the long run. Therefore, we point out the significance of policy implications of real prices and income in the Turkish housing market.

The second essay identifies the macroeconomic factors behind the sovereign credit ratings of global emerging markets assigned by Standard and Poor's (S&P). The financial integration

and globalization of capital markets have facilitated the capital inflows/outflows among countries. Sovereign credit ratings have served as a signal for countries' economic, financial and political situation. Ratings are very important in the sense that they attract capital inflow and investments. This is especially vital for emerging markets. Although the rating agencies do not explicitly reveal their methodologies, it is possible to guess the effects of several variables on ratings by using various econometric models. Concerning the heavy criticisms on rating agencies' performances, we wish to examine the sovereign credit ratings within a specific country-category. In this essay, we study the effects of macroeconomic factors on the sovereign ratings of emerging markets. Using several approaches, we find that the most relevant factors are Budget Balance/GDP, GDP per capita, Governance Indicators and Reserves/GDP. Moreover, our model predicts up to 93% of all credit rating levels. Interestingly, we obtain that S&P's evaluation of the sovereign credit rating for Turkey performs poorly, especially in the highest rating levels.

Lastly, a new correlation coefficient is introduced in the third essay. The correlation in time series has received considerable attention in literature. Its use has attained an important role in the social sciences and finance. For example, pair trading in finance is concerned with the correlation between stock prices, returns, etc. In general, Pearson's correlation coefficient is employed in these areas although it has many underlying assumptions which restrict its use. Here, we introduce a new correlation coefficient which takes into account the lag difference of data points. We investigate the properties of this new correlation coefficient. We demonstrate that it is more appropriate for showing the direction of the covariation of the two variables over time. We also compare the performance of the new correlation coefficient with the Pearson's correlation coefficient and Detrended Cross-Correlation Analysis (DCCA) via simulated examples.

Keywords: Housing Supply, Housing Demand, Cointegration, VECM, Turkey, Sovereign Credit Ratings, Ordered Response Models, Pearson's Correlation Coefficient, DCCA, Stationarity, Non-stationary Time Series

ÖZET

MAKROEKONOMETRİ ÜZERİNE ÜÇ ÇALIŞMA

Yusuf Varlı

Ekonomi Bölümü, Doktora

Tez Danışmanı: Dr. Orhan Erdem

Aralık, 2013

Bu tez, makroekonometri alanındaki üç çalışmadan oluşmaktadır. İlk çalışma, konut piyasasının talep ve arz taraflarını ayrı ayrı modellemektedir. Ampirik kısımda, Ekim 2007 ve Haziran 2012 tarihleri arasında Türkiye konut piyasasının arz ve talep taraflarındaki temel makroekonomik faktörlerin etkilerini belirlemek için vektör hata düzeltme modeli (VECM) tercih edilmiştir. Konut piyasası ile faiz oranı, gayri safi yurtiçi hasıla (GSYH) ve konut fiyatları gibi bazı makroekonomik temel değişkenlerin eşbütünleşik olduğu bulunmuştur. İki eşbütünleşik denklem delili ışığında, Türkiye'deki konut talebi ve arzı üzerindeki değişkenlerin etkisini incelemek için hata düzeltme modeli tahmin edilmiştir. Talep tarafında bağımlı değişken reel mortgage kredi hacmi iken, arz tarafını açıklayan bağımlı değişken ise yapı ruhsatı izin sayısıdır. Bu çalışma, makroekonomik değişkenlerin mortgage kredilerinin ve yapı ruhsatı izin sayısının dinamik davranışları üzerinde farklı etkilere sahip olduğunu ortaya koymaktadır. Ayrıca, yapısal VECM'e dayalı dürtü yanıtı analizi Türkiye'deki konut piyasasının ekonomideki şoklara duyarlı olduğunu göstermektedir. Bu makale aynı zamanda öngörü hata varyans ayrışmasını (FEVD) gösterir ve reel konut fiyatlarının ve kişi başına düşen reel GSYH'nin uzun vadede konut piyasasındaki önemli rolünü belirtir. Bu nedenle, Türkiye'deki konut piyasası içerisinde reel konut fiyatların ve gelire dayalı politika uygulamalarının önemini işaret etmekteyiz.

İkinci çalışma, Standard and Poor's (S&P) tarafından global gelişmekte olan piyasalara verilen ülke kredi derecelerinin arkasında yatan makroekonomik faktörleri tanımlar. Sermaye piyasalarının finansal entegrasyonu ve küreselleşmesi, ülkeler arasındaki sermaye

girişlerini/çıkışlarını kolaylaştırmıştır. Ülke kredi dereceleri ekonomik mali ve siyasi durum için bir sinyal görevi görmektedir. Sermaye girişlerini ve yatırımları çekmek anlamında bu dereceler çok önemlidir. Bu durum, gelişmekte olan pazarlar için hayati derecede önemlidir. Derecelendirme kuruluşları açıkça kendi metodolojilerini ortaya koymamasına rağmen, çeşitli ekonometrik modeller kullanılarak farklı değişkenlerin kredi dereceleri üzerindeki etkileri tahmin edilebilir. Kredi derecelendirme kuruluşlarının performanslarına gelen ağır eleştirilere ilişkin olarak, belirli bir ülke kategorisi içinde devlet kredi derecelerini incelemek istemekteyiz. Bu çalışmada biz, makroekonomik faktörlerin gelişmekte olan piyasaların kredi derecelerine etkilerinin incelemekteyiz. Çeşitli yaklaşımlar kullanılarak, en uygun faktörlerin Bütçe Dengesi/GSYH, kişi başına düşen GSYH, Yönetişim Göstergeleri ve Rezervler/GSYH olduğu bulunmuştur. İlginçtir, S&P'nin Türkiye için kredi notu değerleme performansının, özellikle yüksek kredi derecelerinde, zayıf olduğuna ulaşılmıştır.

Son olarak, üçüncü çalışmada yeni bir korelasyon katsayısı tanıtılmaktadır. Literatürde, zaman serileri içerisindeki korelasyon kavramı oldukça dikkat çekmektedir. Korelasyonun kullanımı, Sosyal Bilimler ve Finans alanlarında önemli bir role ulaşmıştır. Örneğin, Finans'daki "pair trading" işlemi, hisse senedi fiyatları, getirileri vb. değerlerin korelasyonu ile alakalıdır. Genel olarak, kullanımını kısıtlayan birçok temel varsayımları olmasına rağmen, Pearson korelasyon katsayısı yukarıda bahsedilen alanlarda kullanılmaktadır. Burada biz, veri noktalarının gecikme farkını hesaba katan yeni bir korelasyon katsayısı tanıtmaktayız ve bu yeni korelasyon katsayısının özelliklerini incelemekteyiz. İki değişkenin kovaryasyonunun zaman içerisindeki yönünü göstermede, yeni korelasyon katsayısının daha uygun olduğunu göstermekteyiz. Ayrıca, yeni korelasyon katsayısı ile Pearson korelasyon katsayısının ve Eğilimden Arındırılmış Çapraz Korelasyon Analizi'nin (DCCA) performanslarını, simüle örnekler üzerinden karşılaştırmaktayız.

Dizin Terimleri: Konut Arzı, Konut Talebi, Eşbütünleşme, VECM, Türkiye, Devlet Kredi Dereceleri, Sıralı Yanıt Modelleri, Pearson Korelasyon Katsayısı, DCCA, Durağanlık, Durağan Olmayan Zaman Serileri

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks and gratitude to my supervisor Orhan Erdem for his invaluable support. He not only has guided me with this dissertation, but also has led me to be a better person.

With respect and appreciation, I would like to thank all my family who have given me unfailing moral and material support. In particular, special thanks for my mom who has dedicated her life to her two sons. I would also like to express my sincere appreciation to my glorious friends for their encouragement.

My greatest acknowledgments go to my deary wife, Büşra. I would not finish this dissertation without her understanding, cooperation and kindness.

Contents

List of Figures

List of Tables

1 Introduction

The 2007-08 Global Financial Crisis has revealed the need of applied studies and theoretical works which have applications in Economics. The use of applications in both Econometrics and Macroeconomics gathers these two fields and this togetherness is named Macroeconometrics. This thesis contributes to expanding the area of Macroeconometrics by submitting two empirical essays and one theoretical essay which bears a number of applications.

In the first essay, we model both the demand and supply sides of the housing market and present an empirical analysis for Turkey using the vector error correction model (VECM). Both the demand and supply sides in the housing market are modelled in reference to the problem of representative household and firm. Therefore, maximizations of the intertemporal utility functions with respect to budget constraints for both of the agents are the main concerns dealt with in this study. Additionally, although there are many studies related to the housing market in the literature, few studies have been performed in Turkey because of the lack of data on house prices. In this essay, the demand and supply sides of the housing market are modelled using mortgage credit volume as the demand side variable and construction permit numbers as the supply side variable. Its contribution to the literature is the analysis of the effects of macroeconomic factors on the housing credit and the number of construction permits. Concerning the active interventions of the Central Bank of Turkey on credit markets, together with the recently adopted renovation policies by Ministry of Environment and Urbanization, the simultaneous control of interest rates policies and number of construction permits is of crucial importance.

The effects of the macroeconomic factors on the sovereign ratings of emerging markets assigned by Standard and Poor's are examined in the second essay. Although numerous studies have examined the country categories of 50 to 90 countries, few have been performed within country-category studies. The main contribution of this essay to the existing literature is to form a study for the category of emerging markets using quarterly panel data. We use both linear and ordered response type estimations. We also present post estimation re-

sults and explanatory variables' overall performance for emerging markets. In the country specific analysis, the performance of explanatory variables in the model for Turkey is poor when compared to other emerging countries. Keeping in mind that the real GDP of Turkey has doubled, that inflation has decreased from three digits to one digit, and that interest rates have decreased from upwards of 90% to 7% between 1994 and 2010 thereby making the country more attractive to international investors and funds; it would be valuable to determine and discuss the country-specific factors of Turkey's sovereign credit ratings.

Finally, we promote a new correlation coefficient which takes into account the lag difference of data points. Various financial models, such as pairs trading, are concerned with the correlation between two different time series data, e.g., stock prices or returns. Pearson's product moment correlation coefficient is the most commonly used quantity to measure such correlations. However, there are many underlying assumptions (such as stationarity) for the validity of this coefficient. The Formula of Pearson's correlation coefficient takes into account the distance of two variables from their means. In this study, the pitfalls of the Pearson's correlation coefficient have motivated us to introduce a new correlation coefficient that measures the distance between two subsequent data points by taking into consideration lag difference. Although the very first data point is lost, we demonstrate that the new correlation coefficient better captures the direction of the covariation of the two variables over time. We also propose various extensions of this coefficient to obtain more reasonable and reliable results at the expense of having more complex formulas.

2 The Demand and Supply Model of Housing: Evidence from the Turkish Housing Market

2.1 Introduction

Since the housing market is among the fundamental factor which shed light on the whole economy, it is one of the most interesting fields of research. Importance of the housing market is well understood especially after 2007-08 Global financial Crisis. Ellis (2008) examines the US housing market, and compare with other countries. She emphasis on the impact of excess housing supply during the housing boom in US and finds that the housing demand is vulnerable to falling housing prices. For the other countries which have more stable credit conditions Duca et al. (2010) indicates that they were affected from the crisis through international channels. For those countries, any boom in housing market owed more to the demand and supply factors in the market.

Two sides of housing market, demand and supply, interact to determine housing prices. On the other hand, economic activities and prices may affect the demand and supply side of the housing market. The efficacy of the demand side approach depends upon the size of the elasticities of the macroeconomic fundamental factors behind housing demand. In addition to the macroeconomic determinants, the factors of production in the housing industry are one of the leading indicators to determine housing supply. Since it is not possible to distinguish housing market from the whole economy, several studies have been performed and many methods have been employed to examine the housing market.

There are mainly two different approaches to investigate the housing market; to examine house prices and to analyze the demand and supply side of the housing market. Several studies have examined how house prices are affected by changes in macroeconomic fundamentals, such as interest rate, income, etc. The general idea regarding this issue is to add demand and supply side variables to the house prices function. Durmaz (2011) argues that

the most efficient analysis can be obtained when both sides will be added to the analysis. Similarly, Chen and Patel (1998) model house prices dynamics as a function of the supply and demand side variables, such as total household income, house completions, construction cost, short-run interest rate, and stock price index. Abelson et al. (2005) explains the changes in real house prices in Australia, finding that income, unemployment rate, real mortgage rates, equity prices, and consumer price index are elastic in relation to house prices.

Moreover, several other studies have also examined the housing market from the point of supply and demand side analysis. To overcome the uncertainty of income elasticity on housing demand, De Leeuw (1971) examines the relationship between housing demand, income, and house prices by performing a cross sectional analysis. According to the results of his study, income elasticity of housing demand is positive. Additionally Schwab (1983) proposes three different views of the relationship between expected inflation and the demand for housing, concluding that demand is not merely a function of nominal interest rate, but also a function of expected inflation and the real interest rate. Carliner (1973) investigates the income elasticity of housing demand by using a panel regression analysis, finding that all income elasticity estimates are both significant, and are measured between zero and one. Additionally, Glennon (1989) examines the elasticities of income, prices, and interest rate to housing demand by using time series data. He finds that the income elasticity of housing demand is positive whereas the price and interest rate elasticities are negative. Apart from the demand side, there is a large corpus of literature on modeling the housing supply of new homes. Topel and Rosen (1988) argue that current asset prices are sufficient statistics for housing investment if both short-run and long-run investment supplies are the same. If the cost of production is affected by the changes in the level of construction activity, then supply is less elastic in the short run than it is in the long run.

Other research methodologies in used to understand the housing market include cointegration and VECM. On the one hand, there are some studies, like Hofmann (2001), and

Brissimis and Vlasssopoulos (2009), which find one cointegration and cover the demand side of the housing market. Hofmann (2001) analyzes the relationship between bank lending and property prices respectively, both for Hong-Kong and for a set of industrialized countries. He concludes that both long-run and short-run causality goes from property prices to bank credit. According to the results of Brissimis and Vlasssopoulos (2009), a line of causality running from housing loans to housing prices in the long run is not confirmed. In the short run however, the analysis provides a contemporaneous bi-directional dependence between housing loans and housing prices. As the availability of credit increases, the demand for property will also rise. Based on the fixed amount of real estate supply in the short-run, real estate prices will tend to go up.

On the other hand, several studies expand the examination and cover not only the demand side of the housing market, but also housing price dynamics. Gimeno et al. (2006) examines the dynamic interaction between house prices and mortgage credit in Spain in which he identifies two cointegration relationships whose dependent variables are interdependent in the long-run. Another study which is examined by Valverde and Fernandez (2010) covers mortgage credit and house prices, finding that interest rates influence lending and house prices in the same direction, whereas they find that house prices have a negative effect on mortgage credit and that real salary has a negative effect on house prices. Finally, Kenny (1999) and Meese and Wallace (1994) employ cointegration techniques in order to discover the long run relationships among housing fundamentals. Using housing completions as the demand side, real housing prices and nominal mortgage rates are found to be cointegrated. Higher interest rates decrease housing demand whereas they increase prices. Meese and Wallace (1994) concludes that the speed of adjustment after a demand shock in the Paris dwelling market was approximately 30% per month.

Although there are many studies related to the housing market in the literature, few studies deal with Turkey due to a lack of data, especially house price data. Bulut (2009) investigates both demand and supply sides by using cointegration analysis, and although

she finds the expected signs in both sides, the results of her study are controversial for the Turkish housing market because she uses the value of dwellings as a proxy for house prices.

In the present study, we analyze demand and supply sides of the housing market by using mortgage credit volume as the demand side and the number of construction permits as the supply side dependent variables. Our contribution to the literature is the analysis of the effects of macroeconomic factors on housing credit and construction permit numbers. Concerning the active interventions of the Central Bank of Turkey on credit markets, together with recently adopted renovation policies by the Ministry of Environment and Urbanization, the simultaneous control of interest rates policies and construction permit numbers is of crucial importance. The outline of the essay is as follows: Section 2.2 presents the theoretical model, and Section 2.3 outlines the data and econometric methodology. In Section 2.4, we provide estimation results, and finally Section 2.5 contains the conclusion.

2.2 A Housing Demand and Supply Model

The macroeconomic model in this essay is constructed using the standard modelling of housing demand and supply. Following a number of previous works (Poterba (1984), Skaarup and Bodker (2010)), both the demand and supply sides of the housing market are modeled referring to the problem of representative household and firm. Therefore, maximization of the inter-temporal utility functions with respect to budget constraint for both of the agents is the primary concern in both the housing demand and supply model. Detailed information and derivations for both the demand and supply sides are given in the following subsections.

2.2.1 Demand

In demand modeling, a representative household maximizes its expected lifetime utility function comprising housing and non-housing consumption. The household discounts the

future values of the utility function using the discount rate β which is smaller than 1. As such, the problem of household is:

$$\underset{\{c_t, h_t\}}{max} E\sum_{t=0}^{\infty}\beta^t u(c_t,\ h_t) \tag{1}$$

where c_t and h_t stand for the consumption of non-housing (or composite) good and housing good respectively. Also, the household subject to following lifetime budget constraint

$$c_t = y_t - p_t h_t \tag{2}$$

where y_t represents the real income of household and p_t is the real house price while the price of composite good is given as numeriare. The real consumption of housing h_t is given as the multiplication of user cost of housing ω_t and housing stock H_t, i.e., $h_t = \omega_t H_t$. Additionally, the constant risk free rate r is used to discount the future values of the budget constraint and the discount rate for any time t is $(1+r)^{-t}$. For simplicity, the discount rate β of the utility function is taken as $\beta = 1/(1+\delta)$, making the discount rate for the utility function at any given time t to become $(1+\delta)^{-t}$.

Denoting with λ_t the Lagrange multiplier on the budget constraint, the first order conditions with respect to state variables c_t and h_t are:

$$(1+\delta)^{-t}u_c = \lambda_t(1+r)^{-t} \tag{3}$$

$$(1+\delta)^{-t}u_h = \lambda_t p_t(1+r)^{-t} \tag{4}$$

where u_c and u_h refers to marginal utilities. The marginal rate of substitution between composite and housing consumption is reached by dividing first order conditions with each other. It is equal to ratio of prices, i.e.:

$$MRS = \frac{u_h}{u_c} = p_t \tag{5}$$

because the price of composite good is given as numeriare. Moreover, household consumption over the time can be derived by using standard Euler Equations. Dividing first order conditions to their one period ahead first order conditions, the Euler Equations can be found:

$$\frac{u_c(c_t,\ h_t)}{u_c(c_{t+1},\ h_{t+1})} = \frac{(1+r)}{(1+\delta)} \tag{6}$$

$$\frac{u_h(c_t,\ h_t)}{u_h(c_{t+1},\ h_{t+1})} = \frac{(1+r)}{(1+\delta)}(1+\pi) \tag{7}$$

where π is growth rate of real house price, is employed as a constant for simplicity. Furthermore, without loss of generality, constant elasticity of substitution (CES) functional form of the utility is assumed, i.e., $u(c_t,\ h_t) = \frac{c_t^{1-\alpha} + h_t^{1-\alpha}}{1-\alpha}$ where $0 \neq \alpha \leq 1$. Then, with using functional form of the utility, MRS in Equation (5) becomes:

$$MRS = p_t = \left(\frac{c_t}{h_t}\right)^{\alpha} \tag{8}$$

which expresses the house price. Using the fact that $h_t = \omega_t H_t$, the logarithm of the MRS in Equation (8) provides that:

$$log(p_t) = \alpha\left[log(c_t) - log(\omega_t) - log(H_t)\right] \tag{9}$$

With the help of the structure of the CES utility function and by using the Euler Equations (6) and (7), expressions of composite and housing consumptions can be rewritten as $c_t = c_0\left(\frac{1+r}{1+\delta}\right)^{\frac{t}{\alpha}}$ and $h_t = h_0\left(\frac{(1+r)(1+\pi)}{(1+\delta)}\right)^{\frac{t}{\alpha}}$. Additionally, the real house price can be written as, $p_t = p_0(1+\pi)^t$. Then, embedding these consumption and price expressions into the budget constraint yields:

$$c_0 \sum_{t=0}^{\infty}\left(\frac{1+r}{1+\delta}\right)^{\frac{t}{\alpha}}(1+r)^{-t} = \sum_{t=0}^{\infty}(1+r)^{-t}y_t - p_0 h_0 \sum_{t=0}^{\infty}\left(\frac{(1+r)(1+\pi)}{(1+\delta)}\right)^{\frac{t}{\alpha}}(1+\pi)^t(1+r)^{-t} \tag{10}$$

The simpler form of Equation (10) is expressed as:

$$c_0 = \kappa_0(\bar{y}_0 - \mu_0 p_0 h_0) \tag{11}$$

where $\bar{y}_0$ is the present value of future income values, i.e., $\bar{y}_0 = \sum_{t=0}^{\infty}(1+r)^{-t}y_t$. Also, $\kappa_0 = \left[\sum_{t=0}^{\infty}\left(\frac{1+r}{1+\delta}\right)^{\frac{t}{\alpha}}(1+r)^{-t}\right]^{-1}$ and $\mu_0 = \sum_{t=0}^{\infty}\left(\frac{(1+r)^{1-\alpha}(1+\pi)^{1+\alpha}}{(1+\delta)}\right)^{\frac{t}{\alpha}}$. Using log-linearization for the Equation (11), it is approximated that:

$$log(c_0) = log(\kappa_0) + \gamma_{\bar{y}} log(\bar{y}_0) - \gamma_{\mu ph}\left[log(\mu_0) + log(p_0) + log(h_0)\right] \tag{12}$$

where $\gamma_{\bar{y}} = \frac{\bar{y}}{\bar{y} - \mu ph}$ and $\gamma_{\mu ph} = \frac{\mu ph}{\bar{y} - \mu ph}$. Here, the values of $\bar{y}$, μ, p and h are used to identify the steady state values or the values in which an approximation is defined. When Equation (12) is inserted into Equation (9) at time zero, the demand equation is found to be:

$$log(H_0) = \Phi + \Pi_y log(\bar{y}_0) - \Pi_p log(p_0) - \Pi_\omega log(\omega_0) \tag{13}$$

where $\Phi = \frac{log(\kappa_0) - \gamma_{\mu ph} log(\mu_0)}{1 + \gamma_{\mu ph}}$, $\Pi_y = \frac{\gamma_{\bar{y}}}{1 + \gamma_{\mu ph}}$, $\Pi_p = \frac{\frac{1}{\alpha} + \gamma_{\mu ph}}{1 + \gamma_{\mu ph}}$ and Π_ω=1. All the parameters of Π's are positive, so the signs before the parameters specify how the housing stock variable is affected by independent variables in the demand equation. Furthermore, Poterba (1984) indicates that user cost is a function of some variables, such as nominal interest rate and expected inflation. Since nominal interest rate determines the mortgage payments, user cost has a positive relation with the nominal interest rate. Moreover, expected inflation indicates the capital gain amount, showing that user cost has a negative relation with expected inflation. For modeling purposes, the user cost can be defined simply as $\omega_t = (i_t)^{\Pi_i}(\pi_t^e)^{-\Pi_\pi}$, where i_t and π_t^e represent nominal interest rate and expected inflation respectively, and Π_i, $\Pi_\pi > 0$. Therefore, the demand function in Equation (13) becomes:

$$log(H_0) = \Phi + \Pi_y log(\bar{y}_0) - \Pi_p log(p_0) - \Pi_i log(i_0) + \Pi_\pi log(\pi_0^e). \tag{14}$$

The demand equation is constructed upon the dependent variable of housing stock. Real

income has a positive effect on housing stock, while the real house price affects the stock negatively. Thus, the more income individuals have at their disposal, the more demand they have for housing. The individual who has more income wants to buy new houses either to further increase his/her amount of income by taking rent from each additional home or desires to raise his/her living standards. House prices in the equation have a standard negative impact on demand for housing. Additionally, an increase in the nominal interest rate and expected inflation change the demand in negative and positive ways, respectively. Rising nominal interest rate makes housing debt more costly, thereby causing demand to decrease. Inflation reduces the effective cost of home ownership (see Poterba (1984)), thereby leading to an increase in housing demand.

2.2.2 Supply

Modeling the supply side of the housing market is concerned with a representative firm that maximizes its lifetime profits. Under a capital stock model of the housing sector, the representative firm optimizes the investment choice for the maximization of the expected present value of lifetime profits. Therefore, the problem of the firm is:

$$\underset{\{H_t,\, I_t\}}{max}\, \Pi = E\sum_{t=0}^{\infty}(1+r)^{-t}\left[p_t^H H_t - \left(p_t^I I_t + \Phi(H_t,\, I_t)\right)\right] \tag{15}$$

subject to the capital accumulation constraint:

$$H_{t+1} = (1-\delta)H_t + I_t \tag{16}$$

where H_t, I_t and δ stand for housing stock, investment and housing stock depreciation respectively, whereas, p_t^H and p_t^I are the prices for stock and investment variables. Therefore, the amount of return from housing is represented by $p_t^H H_t$ and the cost of the firm is indicated by the sum of the direct investment cost $p_t^I I_t$ and the investment adjustment cost $\Phi(H_t,\, I_t)$. The price of direct investment p_t^I investment is assumed to be formed

by two parts; that is, by real interest rates ri_t and construction cost of investment cc_t. For modelling purposes, the functional form of p_t^I is constructed in the following way: $p_t^I = ri_t^{\alpha} cc_t^{\gamma}$.

The installation of new housing capital involves a convex adjustment cost $\Phi(H_t,\ I_t) = \frac{\psi(\frac{I_t}{H_t} - \delta)^2 H_t}{2\delta}$ which is homogenous of degree one. Furthermore, the adjustment cost is convex in I such that $\Phi'_I > 0$, $\Phi''_I > 0$ and satisfies the condition of t $\Phi'_H < 0$. The latter condition means that the adjustment cost is lower for those firms with more capital stock. Feichtinger et al. (2001) illustrates that "...the fact that when installing a new machine a small firm has to stop production completely while a large firm is more flexible because there production can continue on a parallel production line". Furthermore, ψ in the adjustment cost function indicates the elasticity of housing investment to the price of housing capital stock.

Here, the Lagrange equation is used to solve the maximization problem of the firm. The Lagrange multiplier λ_t on the capital accumulation constraint is redefined considering the shadow price q_t, i.e., $\lambda_t = q_t(1+r)^{-t}$. That is, the price of one extra unit of housing capital for today is equal to the discounted shadow price of housing capital in the next period. Then, the first order conditions with respect to state variables H_t and I_t are:

$$\Phi'_I = q_t - p_t^I \tag{17}$$

$$\Phi'_H = (1-\delta)q_t - (1+r)q_{t-1} - p_t^H \tag{18}$$

Real house price is assumed to have a constant growth rate of π, then evolution of price can be written in as $p_t = p_0(1+\pi)^t$. Using the forward induction method in Equation (18), it can be rewritten in the following way:

$$q_0 = -\frac{1}{(1-\delta)} \sum_{t=1}^{\infty} \left(\frac{(1-\delta)}{(1+r)} \right)^t \Phi'_H + \frac{1}{(1-\delta)} \sum_{t=1}^{\infty} \left(\frac{(1-\delta)(1+\pi)}{(1+r)} \right)^t p_0^H \tag{19}$$

Then, inserting Equation (19) into the time zero value of Equation (17) and yields the

following:

$$\Phi'_I + \frac{1}{(1-\delta)} \sum_{t=1}^{\infty} \left(\frac{(1-\delta)}{(1+r)} \right)^t \Phi'_H = \frac{1}{(1-\delta)} \sum_{t=1}^{\infty} \left(\frac{(1-\delta)(1+\pi)}{(1+r)} \right)^t p_0^H - p_0^I \qquad (20)$$

First derivatives of the adjustment cost function with respect to investment and stock variables at time zero are; $\Phi'_I = \frac{\psi}{2} \left(\frac{I_0}{H_0} - \delta \right)$ and $\Phi'_H = -\frac{\psi}{2\delta} \left[\left(\frac{I_0}{H_0} \right)^2 - \delta^2 \right]$. The variable Φ'_{H} on the left hand side of the Equation (20) normally represents the value of first derivation for all time periods $t \in [1, \infty)$. However, it can be assumed to fixed at time zero due to constancy of the housing stock in the short run. Further explanation is given at the end of this subsection. Therefore, with the use of the first derivatives of adjustment cost and with some amount of simplification, the Equation (20) becomes:

$$\frac{\psi}{2} \left(\frac{I_0}{H_0} - \delta \right) - \frac{\xi_0}{(1-\delta)} \frac{\psi}{2\delta} \left(\left(\frac{I_0}{H_0} \right)^2 - \delta^2 \right) = \frac{\theta_0}{(1-\delta)} p_0^H - p_0^I \qquad (21)$$

where $\xi = \sum_{t=1}^{\infty} \left(\frac{(1-\delta)}{(1+r)} \right)^t$ and $\theta = \sum_{t=1}^{\infty} \left(\frac{(1-\delta)(1+\pi)}{(1+r)} \right)^t$. Both the left and right hand sides of Equation (21) are simplified by using basic log-linearization techniques. Details of simplifications and calculations are presented in Appendix A.

Finally, by using the log-linearization in Equation (21) and remembering that the functional form of direct investment price, i.e., $p_t^I = ri_t^{\alpha} cc_t^{\gamma}$, the representative firm's supply equation with time zero values can be deduced as:

$$log \left(\frac{I_0}{H_0} \right) = \Theta + \Gamma_p log(p_0^H) - \Gamma_{ri} log(ri_0) - \Gamma_{cc} log(cc_0) \qquad (22)$$

All the parameters of Γ's are positive (see Appendix), showing that signs before the parameters specify how the dependent variable will be affected by independent variables in the demand equation. Solomon (2010) argues that the representative firm takes the aggregate housing stock as given when choosing the investment. Moreover, Baer (1986) states that new construction is the main source of existing stock. The model in Brueckner (1981) allows for housing stock to be assumed as being constant over time. A number

of previous works employ a constant housing stock in the short run, using housing starts as a dependent variable in housing supply functions (see Stroebel and Floetotto (2012), Topel and Rosen (1988)). In this manner, the housing stock variable can be embedded in the constant, making the direct investment, i.e., housing starts, able to be taken as the dependent variable of the supply function.

The supply function in Equation (22) indicates that the real house price has a positive effect on housing starts whereas the real interest rate and real construction cost negatively affect new housing. It is held to be standard that an increase in prices has a positive impact on the quantity of new houses supplied. The interpretation of the coefficient of the real interest rate in the supply equation is related with the concept of the opportunity cost of investment. As the real interest rate increases, the opportunity cost of investing also rises, creating a situation in which housing starts are expected to decrease. Lastly, the coefficient of the construction cost variable represents how the cost of building changes the construction market. An increase in the real construction cost makes the investment more costly. Since the investor wants to maximize profits, s/he reduces the level of new housing when construction cost moves up.

2.3 Econometric Methodology

2.3.1 Data

This study is relevant for country specific research. Turkey is chosen in order to understand the relation between macroeconomic fundamentals and the fast growing housing sector. Monthly data between October 2007 and June 2012 is used. The reason for such a limited range is the absence of house price data for Turkey. Data which do not have this frequency are converted using the quadratic matching method. While the logarithm of each datum is used for all calculation, the logarithmic transformation is not employed for neither the real and nominal interest rates nor for inflation variables. Furthermore, those variables

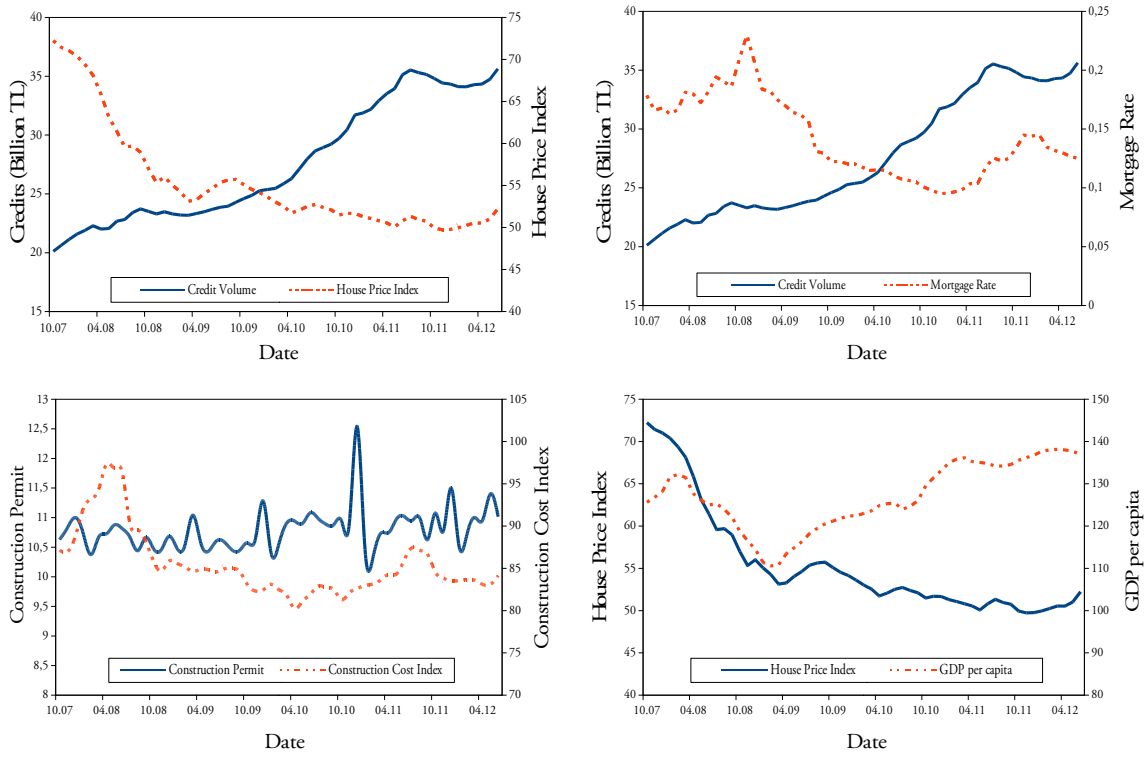
with nominal values are deflated using the Consumer Price Index (CPI), thereby basing the analysis in this essay on the real terms.

Although the demand analysis needs the housing stock variable, housing stock data have not been well recorded in Turkey. However, total mortgage debt constitutes the largest proportion of the housing stock value (see Follain and Dunsky (1997), Poterba (1992)). Thus, for the demand side of the housing market in Turkey, the mortgage credit volume has been selected instead of housing stock. The mortgage credit volume has been taken from the Banking Regulation and the Supervision Agency (BDDK) and then deflated. Construction permit numbers are used as the housing start variable, which forms the supply side dependent variable of the housing market and which has been obtained from the Turkish Statistical Institute (TUIK). The real values of the Gross Domestic Product per capita, which is the proxy for real disposable income, is calculated and announced by TUIK. The nominal interest rate employed in this study is the average mortgage rate. The interest rate and expected inflation data have been taken from the Central Bank of the Republic of Turkey (CBRT). The real interest rate has been derived by subtracting the effect of expected inflation from the average mortgage rate. Furthermore, house price data which have been recently announced by Reidin have been obtained from that company and house prices have been deflated using CPI. Finally, the nominal values of the construction cost index are provided by TUIK which have again been deflated. In addition to this, there is an outlier in the number of construction permits because of a change occurring in the housing permit law during December 2010. The new law brought a number of both extra and costly auditing conditions for the construction sector. As such, a dummy variable is included into the model to investigate the impact of the specific change in the law. Detailed data descriptions and their sources can be found in Table 2.1.

The changes in levels of the variables, which are mentioned in the theoretical model section, varied during the selected time period. The levels of some of the macroeconomic fundamentals in real terms have been illustrated in Figure 2.1. As can be seen from the

graph, the real mortgage credit volume increases dramatically during the selected period whereas the index of the real house price decreases especially during the global financial crisis of 2007. Additionally, the nominal mortgage rate moves down when the real volume of mortgage credit rises. Although the number of construction permits remains relatively level, there is a single outlier due the aforementioned legal adjustment made in December of 2010. There are also several reverse movements in the number of construction permits and construction cost index, particularly after the peak of the global financial crisis. A similar pattern of movements can be seen on the graph which shows the levels of house prices and GDP per capita.

Figure 2.1: Levels of Various Macroeconomic Fundamentals



Note: Each graph illustrates the observations for essential macroeconomic variables during a given time period. All values in each graph are represented by real terms. For convenience, the construction permit variable is scaled logarithmically.

2.3.2 Econometric Method

The OLS estimation method for any model was ruled out since some of the variables in the model are not stationary. When variables in a system are non-stationary (Integrated of order 1, $I(1)$), then the system can be generalized to the multi-variable error correction method which is called VECM (Vector Error Correction Model) and it can be written as:

$$\Delta y_t = \Pi_0 + \ \Gamma y_{t-1} + \Pi_1 \Delta y_{t-1} + ... + \Pi_p \Delta y_{t-p} + \Psi_0 x_t + ... + \Psi_q x_{t-q} + \Phi D_t + \epsilon_t \quad (23)$$

where Π_0 is $(n \times 1)$ vector of intercepts, Π_i's are $(n \times n)$ matrices of short-run coefficients, Γ is $(n \times n)$ structural matrix and ϵ_t is $(n \times 1)$ vector of disturbance terms. Also, x_t and D_t refer to exogenous variables with lag order q and dummy variables, respectively. Since the variables in the system are cointegrated $I(1)$ and Γ has reduced rank (r) so, Γ can be written as $\Gamma = \alpha\beta\prime$ where α is $(n \times r)$ matrix of adjustment coefficients and β is $(n \times r)$ matrix of cointegrating coefficients.

Johansen (1988) develops a procedure to define the cointegrating relation in a multi-variable system. The main purposes of the Johansen procedure are first to determine the number of cointegrating vectors and then to provide the maximum likelihood estimators of determined cointegrating vectors.

Generally, the procedure starts with a test to obtain the order of integration of the variables. All variables must be $I(1)$, because any $I(0)$ variable in the model creates an additional cointegration relationship and the variables which have a different order of integration cause some complications. To determine whether the variables are $I(1)$ or not, unit root tests such as Augmented Dickey Fuller test can be used. The second step is to choose the optimal lag length using an information criteria. The appropriate lag length p can be chosen by minimizing the,

$$SC(p) = \log \det(\widetilde{\mathbf{\Sigma}}_\epsilon(p)) + \frac{\log T}{T} p n^2 \quad (24)$$

where $\widetilde{\Sigma}_\epsilon(p)$ is estimated by $T^{-1}\mathbf{\Sigma}_{t=1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_t\prime$ and $\widehat{\epsilon}_t$ is the estimated residual vector in Equation (23).

Afterwards, one can include deterministic components in the cointegration relations. For different model setups, specific deterministic terms such as intercept and trend place in the model with respect to some economic intuitions.

After the number of cointegrating vectors (r) are estimated using trace test statistics, the conditions for identification of the model must be checked. The identification condition is met when the total number of restrictions (k) is at least the square of the number of cointegrating vectors (r^2). Also, for the previously identified model, the number of restrictions on each cointegrating vector must be equal to the number of cointegrating vectors (Pesaran and Shin (2002)). Restrictions are placed on the Γ matrix. Therefore, both α and β matrices can have required restrictions. The existence of additional restrictions indicate over-identification restrictions on the adjustment coefficients (α) and/or that the cointegrating coefficients (β) can be tested using χ^2 statistics. Restrictions on β matrix form are provided in the linear form. Moreover, restrictions on the adjustment coefficients are imposed if a variable is weakly exogenous meaning that the adjustment coefficient of that variable is not significant.

The general modelling strategy for SVECM is, first, to specify and estimate a reduced form model first and then to focus on the structural parameters and the resulting structural impulse responses.

Another implication of the VECM method in the identification of impulse response functions is called Structural VECM. The strategy of the modelling for structural VECM constitutes two main steps: the first being to specify a reduced form model and the second, to focus on the structural parameters and the structural impulse responses. Lütkepohl (2005)

shows an MA representation of the VECM as:

$$y_t = \Xi \sum_{i=1}^{t} \epsilon_i + \sum_{j=0}^{\infty} \Xi_j^* \epsilon_{t-j} + y_0^* \tag{25}$$

where

$$\Xi = \beta_\perp' [\alpha_\perp' (I_n - \sum_{i=1}^{p} \Pi_i) \beta_\perp]^{-1} \alpha_\perp' \tag{26}$$

and y_0^* contains all initial values.

Then he uses B model for identification purposes where,

$$e_t \sim (0, I_n) \rightarrow \mathbf{\Sigma}_\epsilon = BB' \tag{27}$$

With B matrix, the model has at most r (number of cointegrating vectors) shocks with transitory effects which means that shocks have zero long-run effects. Therefore, there are at least n-r shocks which have permanent effects (Lütkepohl et al., 2004).

Matrix B can be identified if it has n^2 restrictions. Normalization of $\mathbf{\Sigma}_\epsilon$ imposes $n(n+1)/2$ restrictions. The number of transitory shocks (r) provides $r(n-r)$ independent restrictions. For the identification of the permanent shocks, $(n-r)(n-r-1)/2$ additional restrictions are needed and similarly $r(r-1)/2$ additional restrictions must be provided for the identification of transitory shocks.

2.4 Estimation Results

2.4.1 Data Analysis

According to the discussion in the previous section, it is necessary to check whether the variables are integrated of order one or not. Table 2.2 and Table 2.3 show the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test results respectively. According to these tests, the null hypothesis of an existence of unit root cannot be

rejected in levels, but can be rejected in the first differences. We can therefore conclude that all variables are integrated of order one. Here, we suspect a structural break associated with GDP per capita in the first quarter of 2009 in when Turkish GDP performance starts to recover the impacts of 2007-08 global financial crisis. We employ the Zivot-Andrews unit root test to check whether the non-stationarity ($I(1)$) is validated around the break point. According to the test result, we can not reject the null hypothesis which states that GDP per capita variable has a unit root with a structural break in the trend[1]. So, GDP per capita does not become trend stationary with a break. Additionally, it is found that the optimal endogenous lag is four according to the Akaike Info criterion and one in terms of Schwarz criterion. Also, since there is a data scarcity for this study, we prefer to use two lags for endogenous variables in VECM estimation.

[Insert Tables 2.2 and 2.3 here]

2.4.2 Cointegration Test Results

Given the above results, we are able to estimate the VAR model in order to determine the number of cointegrating relationships. The cointegration test proposed by Saikkonen and Lütkepohl (2000) is used in order to determine the number of cointegrating equations. First, this cointegration test estimates the deterministic term, then removes it from the observations, applying a Johansen type test to the adjusted series. The results are reported in Table 2.4. There are two cointegrating equations which explain the long-run relationship between housing sector variables, which were found using Saikkonen and Lütkepohl Cointegration Test results. These two cointegration equations are entitled as the demand and supply equations.

[Insert Table 2.4 here]

[1]According to the results of the Zivot-Andrews test, the chosen break point is January 2009, and lag length is 1. Also, the probability that the null hypothesis can not be rejected is 2,5%. Therefore, we can not reject the null at 99% confidence interval.

According to the discussion in the previous section, in the long run we expect that mortgage credit volume in real term, as the demand side, is related with real house prices, real GDP per capita, nominal mortgage rate and expected inflation. On the other hand, as the supply side, the number of construction permits is affected by real house prices, real construction cost, and real interest rate in the long run. The first cointegrated vector can be attributable to the long run relationship between the demand side of the housing market and the macroeconomic variables, whereas the second one can be attributable for the supply side of the housing market.

As mentioned in the previous section, at least two (number of cointegrating vectors) restrictions have to be placed in each vector. To construct the cointegrating vectors, a number of restrictions must be placed considering the macroeconomic model proposed in Section 2.3. These restrictions can be settled down by the demand and supply functions in Equation (14) and (22) respectively. Thus, the coefficients of construction cost and real interest rate variables are restricted to zero in the demand equation. On the other side, the supply equation indicates that the coefficients of GDP per capita, nominal interest rate, and expected inflation are zero.

The coefficients of the variables (t values are in parentheses) in the cointegrating equations are as follows

$$\begin{aligned} Coint1 \ : \quad cr_{t-1} &= \underset{(26.988)}{12.654} + \underset{(2.494)}{0.004} trend_{t-1} \underset{(-2.864)}{-0.685}\, p_{t-1} + ... \\ &... + \underset{(5.389)}{1.228} y_{t-1} + \underset{(1.759}{2.070} \pi^{e}_{t-1} \underset{(-0.559)}{-0.180}\, i_{t-1} + \end{aligned} \tag{28}$$

$$\begin{aligned} Coint2 \ : \quad per_{t-1} &= \underset{(1.863)}{6.386} + \underset{(3.502)}{0.012} trend_{t-1} + \underset{(3.672)}{2.248} p_{t-1} - ... \\ &... \underset{(-1.081)}{-1.016}\, cc_{t-1} \underset{(-3.297)}{-5.441}\, ri_{t-1} \end{aligned} \tag{29}$$

The cointegration results indicate that all the signs of the coefficients of the selected macroe-

conomic variables on housing demand and supply are same as the signs of coefficients mentioned in the section on the theoretical model. On the demand side (Equation (28)), the real value of mortgage credit is affected by the real GDP per capita and expected inflation positively, and by real house prices and the nominal interest rate negatively. On the other hand, the supply side (Equation (29)) of the housing market states that real house prices impose a positive effect on the level of construction permit, whereas the real construction cost and real interest rate variables affect the variable of construction permit in a negative way. In terms of t values, all coefficients are significant except for the nominal interest rate in the demand equation and the real construction cost in the supply equation. Removing the constant from the equations enables the nominal interest rate variable to become significant. However, there is no differentiation in the model that makes the real construction cost variable significant. Therefore, apart from the variable of real construction cost, all variables in both the demand and supply equations are significant.

The results of the cointegration vector allow us to comment on how the variables are connected in the long run. The aim of the deterministic trend in the cointegration equations is to capture the behavior of trend stationary variables (Kaufman and Cleveland (2001)). As previous studies (Hofmann (2001), Gimeno et al. (2006), Brissimis and Viassopoulos (2009), Valverde and Fernandez (2010)) suggest, mortgage credit volume is normalized and the first cointegrating vector can be debated as the demand side of the housing market. Assuming that the coefficients of the first equation represent the demand side of the equation, the coefficients' signs are consistent with the findings of the macroeconomic model in Section 2.2 and of the work of Kenny (1999) for Ireland. Moreover, several studies have been conducted to ascertain the same relation. Hofmann (2001) examines the relation among bank credit, real price, real GDP, and real interest rate in industrialized countries, concluding that the real GDP has a positive and significant effect on housing demand. Furthermore, Gimeno et. al. (2006), Brissimis and Viassopoulos (2009), Valverde and Fernandez (2010) reach a consensus on the positive relationship between income and housing demand. Moreover, the coefficient signs of the second cointegrating vector, which has

been established as the supply side, are consistent with the supply model proposed in the theoretical model part of this study. The impacts of the variables of house prices, real construction cost, and real interest rate on construction permits are also consistent with the results of Topel and Rosen (1988). The insignificance of the real construction cost variable in the supply function is frequently seen in the literature (see Poterba (1984), Topel and Rosen (1988)). The poor performance of the construction cost estimate in both this study and the existing literature may be caused by a bias in construction cost indexes. Somerville (1999) demonstrates that such a bias is caused by an incorrect measure of labor costs and a failure to address the endogeneity of the costs in the construction market.

2.4.3 Vector Error Correction Model

Error correction model facilitates the understanding of the relationship between the variables in the short run. The coefficients of the speed of adjustment help us to analyze whether or not the short run dynamics converge with the long-run dynamics by following an increasing or decreasing path. Furthermore, they show the speed of convergence and since the demand and supply sides are of concern here, we will only discuss the loading coefficients for the demand and supply side equations.

The Vector Error Correction Models (VECM) are as follows;

$$\Delta cr_t = \underset{(-3.159)}{-0.083}\, coint1_{t-1} - \underset{(-3.510)}{0.009}\, coint2_{t-1} + \underset{(12.690)}{0.665}\, \Delta cr_{t-1} - \underset{(-2.865)}{0.176}\, \Delta p_{t-1} \tag{30}$$

$$\Delta per_t = \underset{(-12.233)}{-1.055}\, coint2_{t-1} - \underset{(-2.281)}{4.811}\, \Delta cr_{t-1} - \underset{(-3.004)}{7.497}\, \Delta y_{t-1} + \underset{(3.040)}{801.953}\Delta \pi^e_{t-1} \ldots$$
$$\ldots - \underset{(-3.002)}{721.241}\Delta i_{t-1} + \underset{(3.010)}{778.184}\Delta ri_{t-1} + \underset{(8.273)}{1.692} dummy_t \tag{31}$$

where the numbers in parenthesis represents t values. In the VECM model, only the variables of mortgage credit volume and construction permit are analyzed in difference

forms. In accordance with the "general-to-specific" approach, only those coefficients of significant variables are shown. In Equation (30), it is observed that there is a short term positive effect in the difference of one lagged mortgage credit volume itself and that the effect of real house prices on credit volume is negative in the short run. These results imply that the short term price elasticity of demand is -0.176. Additionally, a 8.3% of error in credit volume is adjusted or corrected by the demand equation, and a 0.9% of error in credit volume is adjusted or corrected by the supply equation within one period. That is, a disequilibrium in credit volume is adjusted within nearly one year (ten years) by the housing demand (supply). According to Equation (31), it appears that all variables, except real house prices, have a short term effect on construction permits. The performances of the coefficients of the short term effects are poor. This withstanding, the short term income elasticity of supply is found to be -7.497. An error within the construction permit variable is adjusted by the supply equation in a very short term, indicating that the error correction mechanism in the supply of housing is very fast. Furthermore, the VECM residuals are diagnosed for serial correlation, normality, and ARCH, and the results of the diagnostic tests are reported in Table 2.5. The residuals of both mortgage credit and construction permit provide normality, with the diagnostic test results indicating that the model we have used does not suffer serial correlation or ARCH effects.

[Insert Table 2.5 here]

2.4.4 The Structural Model

In this part, we estimate a structural vector error correction model (SVECM) as mentioned in the empirical section. Furthermore, the impulse response analysis is applied to the model by using the estimation results. Our model takes in consideration eight variables, that is $n = 8$. Also, we find 2 cointegration relations in the previous part, $r = 2$. All these information brings up to model as two transitory and four permanent shocks. Therefore, for just identified model, we need $\frac{n(n-1)}{2} = 28$ linear independent restrictions. Because the

cointegration relations are assumed to be stationary, temporary shocks accompanied with two variables: (1) mortgage credit volume and (2) number of construction permits. The restrictions are imposed on the first and second columns of the long-run impact matrix (ΞB). Because, this matrix has reduced rank property, meaning that only $r(n-r) = 12$ restrictions are imposed. Then, for the identification of the permanent shocks, it is necessary to put additional restrictions on $\frac{(n-r)(n-r-1)}{2} = 15$ elements. . It is assumed that the real interest rate and expected inflation variables are determined out of our model, so we put zeros to the sixth and eighth rows of the long run impact matrix. Also, the restrictions for the demand and supply equations in the cointegration analysis are again used here. These assumptions and restrictions enable us to set the associated elements to zero. Furthermore, $\frac{r(r-1)}{2} = 1$ additional restriction is required to identify the transitory shocks. This last restriction takes place on estimated contemporaneous impact matrix (B) because we assume that mortgage credit volume shocks do not push out an immediate effect on construction cost. Hence, the restrictions on the contemporaneous impact matrix (B) and the long-run impact matrix (ΞB) are given as:

$$B = \begin{bmatrix} cr & per & p & y & cc & \pi^e & i & ri \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \\ 0 & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \end{bmatrix} \quad \Xi B = \begin{bmatrix} cr & per & p & y & cc & \pi^e & i & ri \\ 0 & 0 & * & * & 0 & * & * & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 & * \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The imposed restrictions on (B) and (ΞB) help us to estimate our structural vector error correction model, as follows;

$$B = \begin{bmatrix} 0.0221 & -0.0426 & -0.3428 & -0.1448 & 0.1758 & -0.0229 & 0.0270 & 0.0251 \\ 0.7660 & 0.4255 & -8.3239 & 5.1497 & 16.5595 & -4.9655 & 5.5390 & -1.4836 \\ -0.0065 & -0.0036 & 0.6329 & -0.1666 & -0.0870 & -0.1086 & 0.1123 & -0.0845 \\ -0.0139 & 0.0413 & -0.0739 & 0.2814 & -0.1203 & 0.1414 & -0.0902 & 0.0018 \\ 0.0000 & 0.0000 & -0.0250 & -0.3880 & 0.0956 & -0.1154 & 0.1727 & 0.0856 \\ 0.0033 & 0.0018 & -0.1023 & -0.0134 & 0.0253 & 0.0344 & -0.0157 & 0.0282 \\ -0.0160 & 0.0600 & -0.0371 & 0.0415 & -0.0566 & 0.0928 & 0.0019 & 0.0852 \\ -0.0183 & 0.0543 & 0.0410 & 0.0415 & -0.0848 & 0.0623 & 0.0112 & 0.0613 \end{bmatrix}$$

$$\Xi B = \begin{bmatrix} 0.0000 & 0.0000 & -1.0254 & 0.6026 & 0.0000 & 0.2210 & -0.2046 & 0.0000 \\ 0.0000 & 0.0000 & 3.0711 & 0.0000 & -0.4894 & 0.0000 & 0.0000 & -0.4923 \\ 0.0000 & 0.0000 & 1.3471 & -0.2968 & -0.1446 & -0.0883 & 0.1321 & -0.1536 \\ 0.0000 & 0.0000 & -0.0778 & 0.3273 & -0.0793 & 0.1285 & -0.0916 & -0.0877 \\ 0.0000 & 0.0000 & -0.0422 & -0.6563 & 0.1617 & -0.1952 & 0.2920 & 0.1448 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0393 & 0.0152 & 0.0090 & -0.0148 & 0.0094 & -0.0143 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}$$

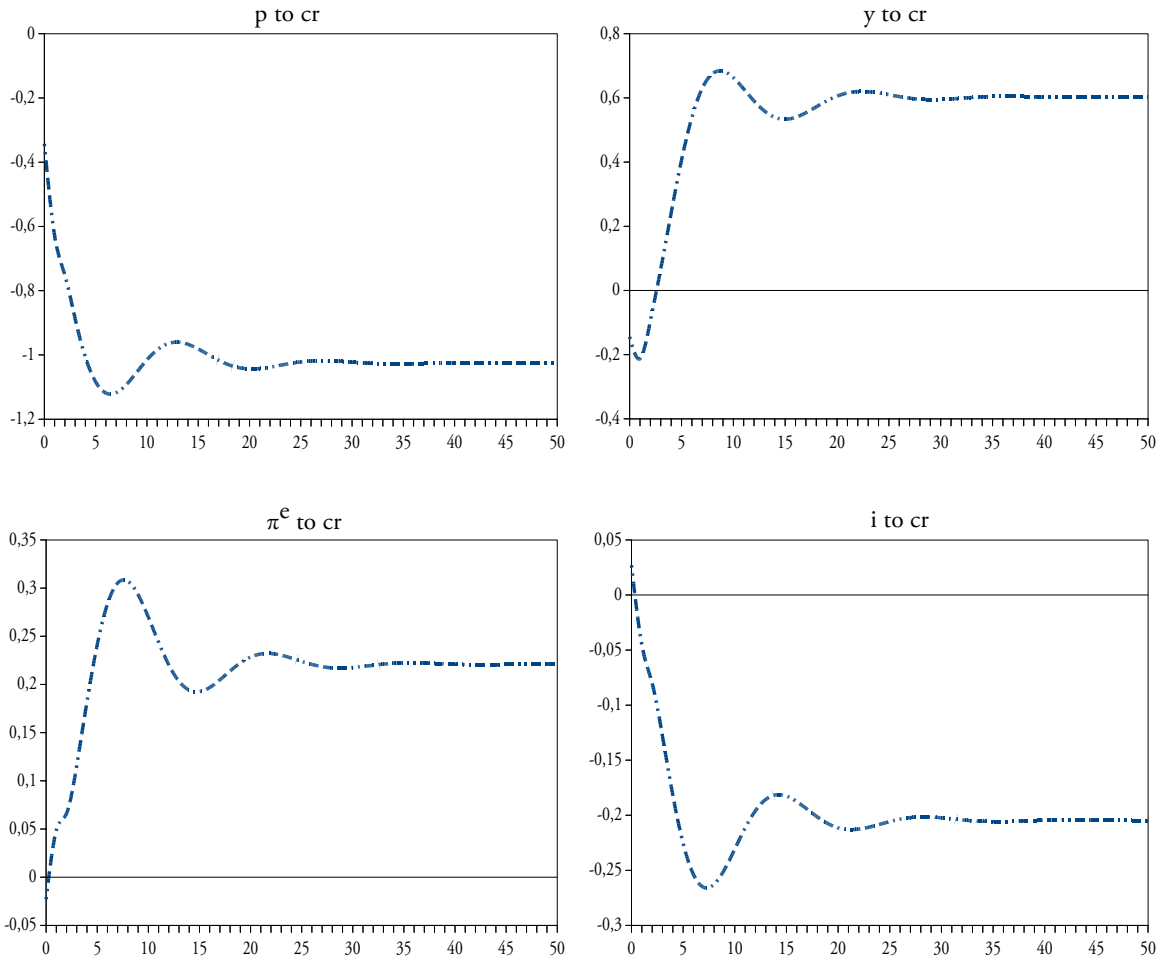
2.4.5 Impulse Response Analysis

An impulse response analysis is used to analyze the interactions between the variables in our model. The figures below demonstrate how a certain variable responds to one standard deviation shock on each variable. In our analysis, we focus on the responses of two variables: mortgage credit volume and construction permit to the impulses on all variables. Since shocks on mortgage credit volume and permit have transitory effects, the impulse responses of mortgage credit volume and permit die out after a certain period. In this section, firstly the demand side shocks and responses will be investigated, afterwhich a supply side analysis will be shown.

The responses of the mortgage credit volume to shocks on the demand side's macroeconomic variables of house prices, GDP per capita, expected inflation, and nominal interest rate can be seen in Figure 2.2. It is demonstrated that the house prices shock affects the mortgage credit volume negatively both in the short run and the long run. There is evidence for the escape behavior of people from the credit market when a shock occurs in house prices. This type of behavior is not very effective in the very short run, but its effect increases with time. At the end, a shock on house prices causes mortgage credit volume to decrease permanently in the long run.

Figure 2.2 also demonstrates how a shock on GDP per capita shock affects the mortgage credit volume. The response is negative in the very short period,. It yields that people tend to consume more non-housing (or non-durable) goods initially while delaying housing consumption when their wealth increase. Therefore, the response of mortgage credit begins with a negative value, becoming positive after three periods, continuing to increase in the medium run. It reaches its permanent level after nearly twenty five periods (two years). The results seem intuitive since a positive shock on the level of income may lead individuals to save more thereby boosting the credit market.

Figure 2.2: Response of Mortgage Credit Volume to Various Shocks on the Housing Demand Variables



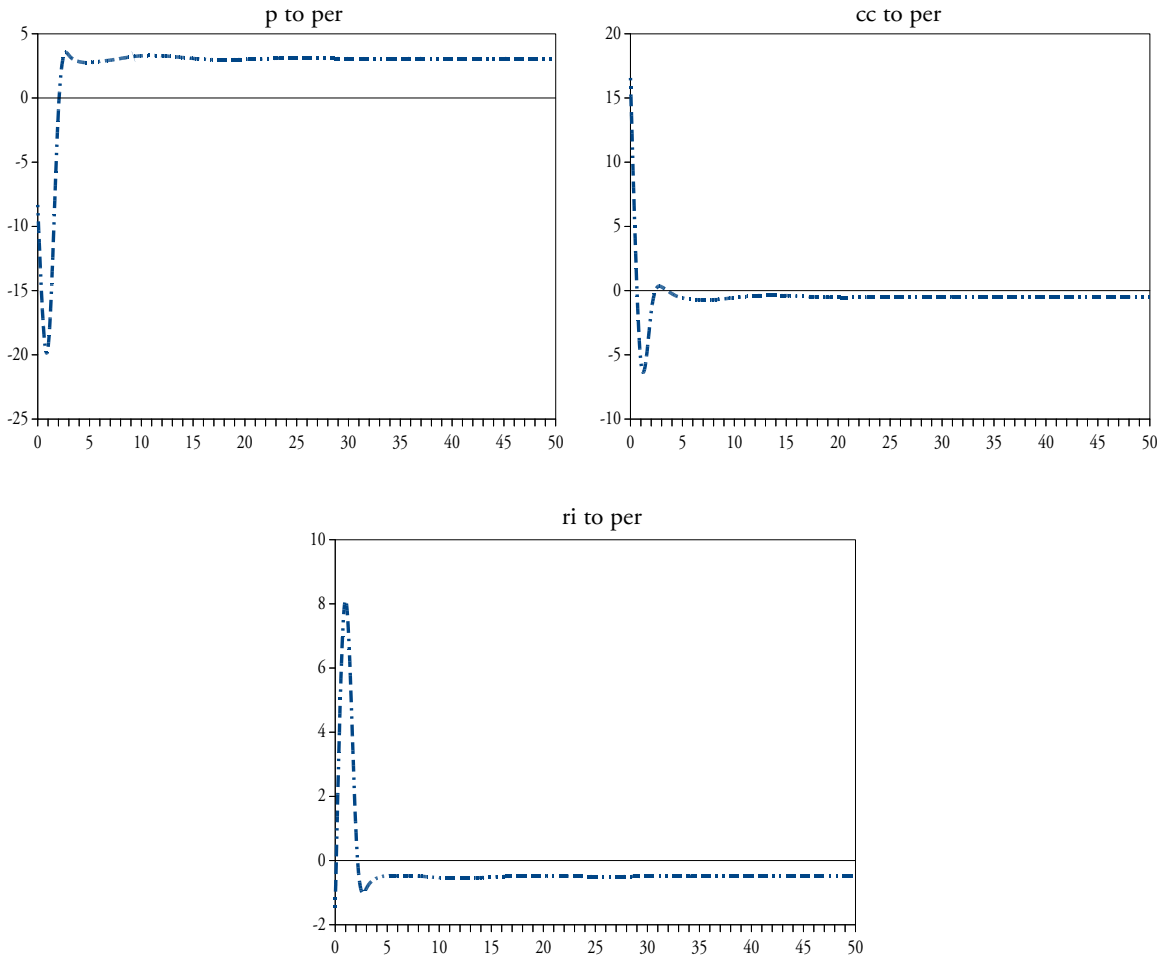
Note: This figure illustrates the responses of the mortgage credit volume variable to the shocks on the demand side macroeconomic variables of house prices, GDP per capita, expected inflation and nominal interest rate. For each graphs, x and y axes represent the time period and the level of response respectively.

The response of mortgage credit to the financial variables; specifically to the expected inflation and the nominal interest rate, is also illustrated in the figure. A positive shock on the expected inflation rate brings a positive permanent response to mortgage credit volume. This result is consistent with the argument that expected inflation reduces the effective cost of home ownership. The mortgage credit volume is also affected by one standard deviation shock on the nominal interest rate. This negative response osculates

in the short and medium run, reaching its permanent level after thirty periods. Mortgage credit becomes more costly with an increasing mortgage rate, so the response of mortgage credit volume to interest rate shock is expected to go through a negative level in the long run.

After the demand side effects of shocks are analyzed, just how the supply side of the model is affected by the shocks will be analyzed. The responses of construction permit numbers to the shocks on the supply side independent variables, such as house prices, construction cost, and real interest rate, are demonstrated in Figure 2.3. The initial response of construction permit to the house price shock is negative. This response may appear due to a misperception of the construction market. Constructors may perceive this to be a shock on general price levels, and may therefore reduce construction activities. After only two periods, the response starts to increase and the effect of the house price shock on construction permit becomes permanently positive in the long run. This is to say that a positive shock on house prices increases the housing starts. The positive relationship between house prices and construction permit in our model supports the general supply relation between prices and quantity.

Figure 2.3: Response of Construction Permit to Various Shocks on the Housing Supply Variables



Note: This figure illustrates the responses of the construction permit variable to the shocks on the supply side macroeconomic variables of house prices, construction cost and real interest rate. For each graphs, x and y axes represent the time period and the level of response respectively.

The next step is to examine how a positive shock on the construction cost index affects the construction permit. It can be seen that a shock on the construction cost index leads to an initial increase in construction permit numbers for a very short run. This can happen due to a misunderstanding by construction suppliers. Only one period later, the effect reverses, becoming negative after three periods. Eventually, the negative effect of the construction cost shock reaches its permanent level after only six periods. Therefore, one

standard deviation shock on the construction cost index induces a decrease in the number of construction permits in both the medium and long run.

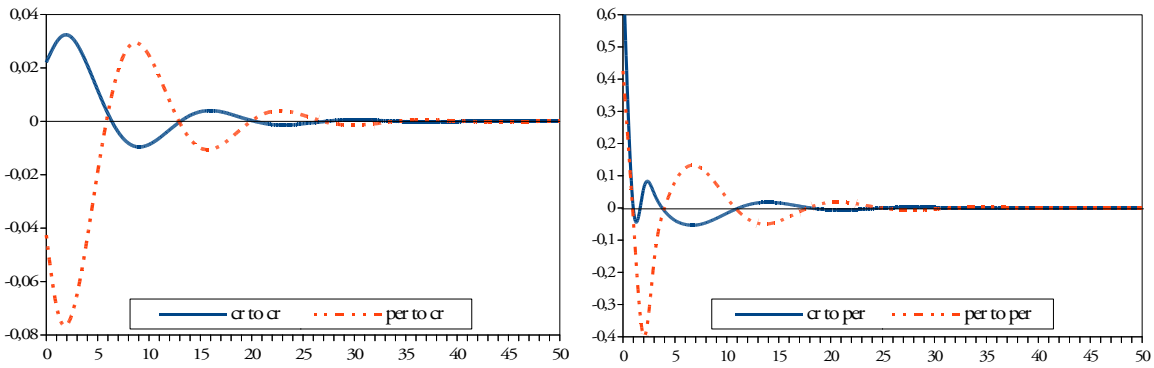
According to the estimation results of the long run impact matrix in the previous section, it is ascertained that a real interest rate shock exerts a negative effect on the number of construction permits in the long run. Figure 2.3 displays how construction permit is influenced by a shock on the real interest rate. The response of the interest rate shock fluctuates at very high levels in the short run, after which fluctuations disappear in the middle run. Here it can be said that the response of suppliers or the construction market to an increase in the real interest rate is unstable until it reaches long run equilibrium. An interest rate shock also exerts an impact on the house prices and, as can be seen in the long run impact matrix, is negative. Therefore, this impact may explain the response of construction permit numbers to the shock on the real interest rate in the long run.

[Insert Table 2.6 here]

Another implication of the impulse responses is the need to investigate how the responses of a macroeconomic variable to shocks on other variables are correlated. The responses of the dependent variables of demand and supply equations to the shocks on the other variables are represented in Table 2.6. The correlation values in the table are listed over different time horizons. Here, it is observed that the responses to the demand side shocks are highly correlated and that the signs of the correlation values are consistent with the respective signs of the variables in the demand equation. However, the correlation values of the responses to the supply side shocks are weak, most likely due to the insignificance of the construction cost variable. Here, the signs of the correlations do not illustrate how the variables affect the overall system in the supply equation. Furthermore, the temporary responses of the shocks accompanied with mortgage credit volume and construction permit—the dependent variables of cointegrating demand and supply equations—are illustrated in Figure 2.4. The first graph demonstrates the response of housing demand to the shock on the supply side and the demand side by itself. The second graph shows how

the shocks on the dependent variables of housing demand and supply affect the supply of the model. Both graphs in Figure 2.4 indicate that all the responses to the shocks on the dependent variables of the model osculate in both the short and medium run. In addition, these shocks lose their influences after nearly 25 periods (2 years).

Figure 2.4: Temporary Responses to the Housing Demand and Supply Shocks



Note: This figure illustrates the temporary responses of the dependent variables of cointegrating demand and supply equations to their accompanied shocks. For each graphs, x and y axes represent the time period and the level of response respectively.

2.4.6 Forecast Variance Decomposition

One of the implications of the structural vector error correction model is variance decompositions. The forecast variance decompositions are summarized in Table 2.7. Here, some horizons are skipped due to small changes in them and 1st, 5th, 20th and 50th horizons are reported as benchmark time measures. It is noted that there is no observable difference between the factors that explain forecast error in the nominal and real interest rates. While the forecast error variance of both house prices and expected inflation are mostly caused by shocks on house prices, the forecast error in the variables of GDP per capita and construction cost are mainly accounted for by GDP per capita in both the short and long runs. Additionally, house prices shocks plays an important role for the dependent variables in the demand and supply analysis. The forecast error variance of the mortgage credit volume is dominated by the shocks on house prices in each time period. The vari-

ability in housing demand can also be explained by the income in the long run. Also, the construction cost shock is the main source of the forecast error variability for the construction permit in the short run, whereas house price shock is the main determinant of the forecast error variability for the permit variable in the long run.

[Insert Table 2.6 here]

2.5 Concluding Remarks

The main purpose of this essay is to analyze the housing market from both demand and supply sides. Both of the demand and supply of housing are separately modelled using the equations in the standard partial equilibrium concept. The maximization problem with respect to budget constraints for both the representative consumer and firm is the main concern while modeling. The housing market model in this study is supported with an empirical analysis. The methods used to analyze the market are a cointegration analysis and the vector error correction model. Even if studies related to this topic exist, this study stands out in the sense that it uses the appropriate data instead of using proxies. Furthermore, the impulse responses and variance decomposition after the VECM are employed to analyze the behavior of the market more clearly. According to the housing market model and empirical analysis, four main results are obtained; namely, that:

1. The mortgage credit volume is used as a dependent variable of the demand side and is affected by real house prices, real GDP per capita, expected inflation, and nominal interest rate. On the supply side, it is found that the dependent variable of construction permit has a long-run relation with real house prices, real construction cost, and real interest rate.

2. In terms of the significant loading factors in the VECM, there exist the adjustments of the real mortgage credit volume and construction permit to the disequilibrium of the housing demand market and adjustment of construction permit to the disequilibrium of the housing supply market.

3. According to the impulse response analysis, it is concluded that the housing market in Turkey is sensitive to shocks in the economy. That is, the fundamentals of both the demand and supply sides of the Turkish housing market have various effects in the short run, and that these fundamentals show the long run impacts which are determined in the Structural VECM.

4. In consideration of the analysis of variance decomposition, our essay reaches a policy implication result stating that, overall, the real prices and real income play the most important roles in explaining the macroeconomic variables affecting the housing market in Turkey.

3 Understanding the Sovereign Credit Ratings of Emerging Markets

3.1 Introduction

Sovereign credit rating is an assessment allocated by rating agencies regarding financial and economic obligations of a specific country. It plays an important role in determining countries' access to international debt markets and the terms of that access. The credit rating evaluation for a country corresponds to the economic, financial, and political performance of that country from the point of capacity in payment of its loans.

The three main rating agencies; namely, Standard and Poor's (S&P), Moody's, and Fitch, use a combination of several quantitative and qualitative variables (economic, social, and political) in order to assign a credit rating to a debtor. Because they do not announce their methodologies explicitly, an important issue is to identify the factors behind their assignment of sovereign credit ratings.

Various studies try to identify and model these factors. These studies can be classified according to either the methodology (OLS analysis vs. ordered response model) or the data type (panel data vs. cross-section data) used. Cantor and Packer (1996) is one of the first studies attempting to assess the relevance of eight important macroeconomic variables explaining rating assignments by Moody's and S&P's. This study is a cross section analysis of 49 countries, both developed and developing, as of September 29, 1995. Using an OLS analysis, they establish the importance of six factors; namely, per capita income, GDP growth, inflation, external debt, level of economic development, and default history, in determining a country's ratings. In a later paper, using the same econometric framework for 49 developing countries with ratings of B- and higher, Rowland (2004) finds that GDP per capita, the economic growth rate, the inflation rate, external-debt ratios, debt-service ratios, the level of international reserves, and the openness of the economy are important

factors of sovereign credit ratings. GDP per capita and debt ratio are significant in both studies, even if Cantor and Packer (1996) also included developed countries in their data set.

By comparing his results with those of Cantor and Packer (1996), Rowland (2004) claims that the rating agencies use a similar framework when evaluating both developed and developing countries. Hisarciklilar et al. (2011) examine the main factors that determine the credit rating using an ordered probit model for both developed and developing countries. They use Moody's data which belongs to 93 countries between the years 1999 and 2010. Their model includes not only macroeconomic variables, but also additional dummy variables for the income level of countries. The results of their study suggest that, holding all other factors constant, low income countries tend to receive lower rating than high income countries, and that this distinction is clearer for the OECD countries than it is for non-OECD countries.

Afonso (2003), however, discovers that the macroeconomic factors affecting the ratings are different in developing and developed countries. Studying the key factors determining sovereign debt ratings using data from 81 developed and developing countries, based on ratings assigned by Moody's and S&P's in June 2001 using a cross-section data and employing both a linear and a non-linear transformation of rating levels, he shows that six variables are the most relevant factors in explaining the credit ratings: (1) GDP per capita, (2) external debt-to-exports ratio, (3) level of economic development, (4) default history, (5) real growth rate, and (6) inflation rate. He concludes that GDP per capita emerges as virtually the sole relevant economic variable for developed countries, while external debt is important for developing countries.

Later studies are more comprehensive in the sense that they use panel data: Afonso et al. (2007) empirically examines the economic determinants of sovereign credit ratings for the period 1995–2005, employing panel estimation and ordered probit approaches with random effects. They find that GDP per capita, real GDP growth, government debt,

government effectiveness, external debt and external reserves, sovereign default indicator, as well as being a member of the European Union are the most important determinants of the sovereign debt ratings. Their evidence also shows that the external debt and external reserves are more relevant for low-rated countries, while inflation played a more profound role for high-rated countries. The importance of external debt for developing countries is in line with our finding in section 3.4. Analyzing a panel data of 16 emerging countries, Rowland and Torres (2004) finds that economic growth rate, the debt-to-GDP ratio, the reserves-to-GDP ratio, and the debt-to-exports ratio are the most important economic determinants of sovereign credit ratings. It should be noted that GDP per capita does not seem to be an important determinant in the study mentioned.

Bissoondoyal-Bheenick (2005) presents an extensive analysis in terms of the number of countries (95 developing and developed), time period (from December 1995 to December 1999), and methodology (an ordered response model). He attempts to analyze the determinants of sovereign ratings separately for high-rated and low-rated countries. The estimated results reveal that current economic and financial indicators are only a part of the input, and that they alone do not determine ratings. Moreover, GNP per capita and inflation seem to be the most relevant economic variables. He also adds that the relevance of economic variables is not the same across the different rating categories. For emerging countries, he says that "... a large range of indicators in the external debt and balance of payments areas are relevant to assign ratings..." In a later study, Bissoondoyal-Bheenick et al. (2006) find that a proxy for technological development is the most significant variable in explaining the ratings assigned to countries.

In the other direction, changes in sovereign credit ratings affect the economy. A number of studies put emphasis on these effects. Specifically, Kaminsky and Schmukler (2002) find out that changes in sovereign debt ratings and outlooks affect financial markets, with average yield spreads increasing 2 percentage points in response to a downgrade. Jaramillo and Tejada (2011) find that investment grade status reduces spreads by 36 percent, above

and beyond what is implied by macroeconomic fundamentals. Using a panel data analysis and event studies, Larraín et al. (2011) present that changes in credit rating have a significant effect on international financial markets. Their study reveals a highly significant announcement effect when developing countries' sovereign bonds are put on review with a negative outlook.

Even though there are numerous studies examining the country categories of 50 to 90 countries, there are not many within country-category studies. The main contribution of this essay to the existing literature is to form a study for the category of emerging markets using quarterly data. After looking at the overall performance of predictions, we also make country specific analysis and compare the performance of countries' credit rating predictions.

The main and private credit risk rating agencies are Moody's, Standard & Poor's (S&P), and Fitch. In this study, we use S&P's ratings, since it is the first agency to being assessing Turkey's performance since 1992, thus presenting us the opportunity to have a bigger data set. Additionally, there are often claims of governments of emerging markets regarding both unfair and biased rating decisions by S&P.

Our data selection is built on the data of 8 countries considered to be emerging markets according to the following explanation: There are a plethora of formed lists of economies in the category of emerging markets. Some of the lists presented in Wikipedia are: the International Monetary Fund list, Columbia University's EMGP (Emerging Market Global Players) List, the FTSE Group list, MSCI Inc.'s list, Standard and Poor's list, the Dow Jones list, the Frontier Strategy Group (F10) list, the BBVA Research list, and the Emerging Markets Index compiled by Mastercard. We have employed the emerging markets category in this study by selecting countries appearing in at least seven of the eight lists; these countries being Brazil, China, India, Indonesia, Mexico, Russia, South Africa, and Turkey.

The global financial crisis of 2007-08 makes it necessary to investigate credit rating agencies and their evaluation methodology. In this essay we try to identify the macroeconomic

factors and indicators behind the sovereign ratings of emerging markets. For this purpose, quarterly data from 2002 to 2011 is selected for 8 emerging markets. We use both linear and ordered response analyses and perform respective estimations. According to our results we find that the following variables, Budget Balance/GDP, GDP per capita, Governance Indicators, and Reserves/GDP are the main determinants of credit ratings for emerging markets. We are also interested in post estimation results, and our models correctly predict up to 93% of all credit rating levels. When we compare our results for Turkey and the other emerging markets employed in this study, we obtain that the explanatory variables' overall performance for Turkey is not as good as it is for the rest of the emerging countries. We therefore came to the conclusion that S&P's evaluation of the sovereign credit rating for Turkey has poor performance, especially in the highest rating levels. Therefore, it seems that the rating agencies deserve to be studied and inspected carefully.

The organization of the essay is as follows: The next section introduces the econometric model which can be separated into linear and ordered response frameworks. Section 3.3 describes the data and rating system of S&P. In Section 3.4, estimation results are presented, and Section 3.5 shows post-estimation results. Finally, Section 3.6 covers the conclusion.

3.2 Econometric Model

3.2.1 Linear Framework

In order to model the ratings of selected emerging markets, we start with the following panel linear regression:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it} \tag{32}$$

where y_{it} represents emerging markets' ratings and x_{it} represents the vector of macroeconomic explanatory variables. Here, i is the index of the "country" and the index t represents the "time period" in the model. Also, β is a vector of coefficients, α_i stands for the individual error terms for each country and the disturbance term ε_{it} is assumed to be

independent across countries and periods.

A simple OLS approach to this analysis has two important caveats, one theoretical, and one empirical. The theoretical problem is related with the construction of the ratings variable and its usage in OLS. An OLS technique assumes that the ratings have equally spaced intervals, which means that upgrading from, say, level 1 to level 2 requires the same effort with an upgrade from level 5 to level 6. This caveat can be remedied by using an ordered response model which we have preformed and whose results are detailed in Section 3.3. The empirical problem is that a rating may not always represent the current situation in a specific country and that its assessment may sometimes take months following a watch-up period. Therefore, a rating may represent some data from the past, That is, error terms of the simple OLS model in Equation (32) are serially correlated. In this case, using an autoregressive (AR) structure for the error terms would prove useful.

In order to test whether or not there is a serial correlation in error terms, we use the Wooldridge test for serial correlation in panel data models (Wooldridge (2002)). This method uses the residuals from a regression of first-differenced variables.

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + \varepsilon_{it} - \varepsilon_{it-1} \tag{33}$$

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \varepsilon_{it} \tag{34}$$

The null hypothesis assumes no serial correlation. As such, the residuals from the regression of first-differenced variables should have an autocorrelation of -0.5, that is $Corr(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}) = -0.5$. A Wald test can be used to test this hypothesis.

If it is found that no serial correlation structure exists, then the simple pooled OLS approach may be employed. However, in the case of autocorrelation in error terms, the best option is to use the pooled OLS with panel-corrected standard errors. It is assumed that,

there is first-order autocorrelation AR(1) within panels in a such way:

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it} \tag{35}$$

where ρ is a coefficient of autocorrelation and u_{it} is independent and identically distributed (i.i.d.) with zero mean and variance σ_u^2.

Besides the approach of pooled OLS, the panel specific effects are analyzed with fixed and random effects linear models using an AR(1) disturbance term. If the country specific error termsα_i's are assumed as fixed, then the model is called fixed effects model. On the other hand, if α_i's have i.i.d. properties or the country specific errors have no correlation with explanatory variables $\mathrm{E}(\alpha_i|x_{it}) = 0$, this model is called as random effects model. The Hausman specification test can be used to determine whether fixed effects model or random effects model is preferable in panel data. Under the null hypothesis, the random effects model is chosen due to its efficiency. Alternatively, the random effects model becomes inconsistent, whereas the fixed effects model is consistent. Therefore, the fixed effects model is preferred in this case.

3.2.2 Ordered Response Framework

The basic OLS technique assumes that the intervals between ratings are equally spaced. For example, an upgrading from level 1 to level 2 requires the same change in variables with an upgrading from, say, level 5 to level 6. This situation may be solved by using an ordered response model. Considering that the dependent variable is restricted to ordinal values with a unit change in between, a nonlinear transformation and an ordered response model would better fit the data. Therefore we would like to construct the following ordered probit model:

$$y_{it}^* = \beta x_{it} + \alpha_i + \varepsilon_{it} \tag{36}$$

where y^*_{it} is an observable variable with certain thresholds, μ_i. If we assume a particular probability distribution for y^*_{it}, we will have the ordered probit model and can calculate the following μ_i's:

$$y_{it} = \begin{cases} 1 & if \quad y^*_{it} \leq \mu_1 \\ 2 & if \quad \mu_1 < y^*_{it} \leq \mu_2 \\ 3 & if \quad \mu_2 < y^*_{it} \leq \mu_3 \\ \vdots & \quad \vdots \\ N & if \quad \mu_{N-1} < y^*_{it} \end{cases} \tag{37}$$

Here N is the highest number of the ordinal values. Therefore N=46, 30, 16 since the highest rating Turkey received from S&P corresponds to 46, 30, 16 in rating Scale 1, 2 and 3 respectively. Hence the number of thresholds change accordingly.

Afonso et al. (2007) mentions that the usage of ordered probit is not clear in panel data structure. We also account the serial correlation in error terms, thereby making the estimation of this model becomes more complex. As Wooldridge (2002) presents, we propose two approaches in here. One is to employ a serially correlated error term in the context of pooled data. Here, normal procedure of the ordered probit is applied but the cluster robust covariance matrix is supposed to be used. The second approach is to use the ordered probit considering random effects. Log likelihood maximization is used here and again the cluster robust standard errors are regarded.

3.3 Data

3.3.1 Credit Rating Schema and Scales

In general, the ratings are variations of the scale A, B, C, or D. Conventionally, the AAA is the top rating given by all agencies. The lower the rating indicates the higher probability of default. Governments rated above “BBB” are considered to have “investment grades”,

while those rated below "BBB" are considered to have "speculative grades". In order to differentiate between governments in the same category, both S&P and Fitch adopt arithmetical symbols (+ and -).

For each government assessed, the agencies publish their findings on the probable direction that the risk rating will take over the medium term (one to three years). This indicator is known as an outlook which can be positive, negative, stable, or developing. When the possibility of change emerges in the sovereign risk rating of a particular government, the agencies may place it on a separate list. S&P calls it a Rating Alert, and such an operation is known as Credit Watch, whose ratings are 'positive', 'negative' or 'stable'.

[Insert Tables 3.1 and 3.2 here]

Table 3.1 summarizes the ratings of S&P and their linear transformation. Because the countries that we selected have received ratings ranging from CCC to AA- and the default rating, we only use the ratings mentioned. We use three types of scaling: Scale 1, 2, and 3. Scale 1 has the largest variation and is based on the one to one transformation considering the Credit Watch. Scale 2 is formed by overlooking the credit watch values in very high and low grades, smoothing the values in Scale 1. Moreover, the letter grades are transformed directly, and this transformation is presented in Scale 3.

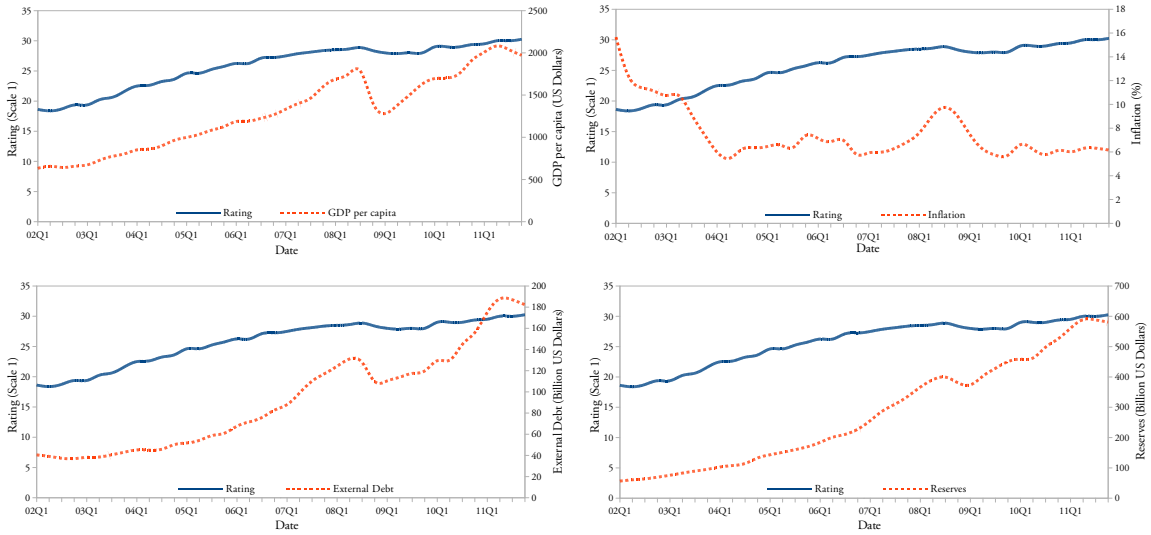
The credit rating migration information of S&P is illustrated in Table 3.2. The transition of credit ratings from one level to another in one period (a quarter of a year) is taken into consideration in order to analyze the general assessment of S&P for emerging markets. Downgrades and upgrades are leveled in notches accordingly. The results show that S&P's willingness to change the ratings for emerging markets. The percentage of no migration in all transitions is between 77% - 88% with respect to the rating scale. For the remaining transitions, the number of rating upgrades is higher than downgrades in all rating scales.

3.3.2 Macroeconomic Variables

The data of key macroeconomic fundamentals which we employ as the determinants of sovereign credit ratings of selected emerging markets are composed of the following variables (The expected signs of their effects on the rating are given in parentheses):

The first variable is "Budget Balance/GDP" (+), general government budget balance value is divided by nominal GDP. The variable states for the income is "GDP per capita" (+); that is, the per capita nominal gross domestic product in US dollars. Population, which is used for determining per capita income, is based on yearly data and is extrapolated into quarterly data. Also, the quarterly data of GDP per capita is seasonally adjusted. Another variable in our model is the "Current Account Deficit/GDP" (-), in which the value of the current account deficit is divided by nominal GDP and is then seasonally adjusted. "Inflation" (-) is the variable which refers to the annual percentage change of consumer prices. The "Governance Indicators" (+) variable is the average level of a number of governance indicators, such as Control of Corruption, Political Stability, and Absence of Violence. It is composed of yearly data and converted to quarterly data. The debt positions of countries are represented by "External Debt/Export" (-), in which the value of total consolidated debt positions in US dollars is divided by the levels of export. "Reserves/GDP" (+) stands for the US dollars amounts of reserves, excluding gold, over the value of the gross domestic product. Furthermore, the "Unemployment" (-) variable corresponds to the seasonally adjusted rate of unemployment. The growth performances of countries are measured by "GDP Growth" (+); that is, the growth rate of real GDP. Finally, we control the currency movements by "Exchange Rate" (-), augmentation of the level of National Currency per U.S. Dollar.

Figure 3.1: Credit Ratings Average and Essential Macroeconomic Variables



Note: Each graph illustrates the observations for the average of the ratings that selected emerging markets received from S&P and the essential macroeconomic variables in a given time period.

The relationships between the average of the ratings that selected emerging markets received from S&P and the average values of some macroeconomic variables in a given time period are shown in Figure 3.1. The effects of the global financial crisis of 2007-08 on the macroeconomic structure of emerging markets can be easily seen from the figure.

3.3.3 Data Description

We use a quarterly data of the credit ratings and key macroeconomic indicators from 2002 to 2011. While nearly all the studies in the literature use yearly data, we collect quarterly data and get them from some sources, namely International Monetary Fund (IMF), World Bank, Organization for Economic Co-operation and Development (OECD) and Bloomberg. Detailed data descriptions and their sources can be found in Table 3.3. Additionally, the emerging countries that we select for this study are Brazil, China, India, Indonesia, Mexico, Russia, South Africa and Turkey.

3.4 Estimation Results

Estimation results are separated (as in Section 2) into linear and ordered response frameworks. All of the analyses are performed using Stata software. The respective codes and commands are used for each estimation. The main parts of the estimation tools and results are presented here. The detailed results, codes etc. are available up on request.

3.4.1 Results of Linear Analysis

Here, the results represent the outcome of the linear analysis using panel data. First of all, the findings for pooled OLS estimation are demonstrated in Table 3.4. There are insignificant variables found for each rating scale. From now on, all estimation analyses for the different rating scales do not include these insignificant variables. Looking at Table 3.4, we see that the Unemployment and Exchange Rate are insignificant variables for all regressions with respect to different scales. Also, the GDP Growth variable is not significant for the analysis of Scale 1 and Scale 2. All significant variables, except GDP Growth in Scale 3 regression exhibits signs which are in line with what we expect. While R^2's are very high for all regressions, at the same time, the Wooldridge test results indicate that all regressions suffer from first order autocorrelation in error terms.

[Insert Tables 3.4, 3.5 and 3.6 here]

In order to deal with the autocorrelation structure in error terms, we re-estimate pooled OLS regression using the panel corrected standard errors. We specify the first order autocorrelation for the pooled OLS and present the results in Table 3.5. The Current Account/GDP variable for all rating scales and the GDP Growth variable for Scale 3 now become insignificant. The remaining variables are still significant as before, showing the expected signs. Furthermore, we can say that considering the autocorrelation structure in error terms causes lower R2's for all rating scales. Table 3.5 also gives the estimated

autocorrelation parameter ρ that represents the impact of serial correlation in error terms. In other words, the high estimated values of parameter ρ for each rating scale indicates the S&P's conservator behavior to change the ratings for emerging markets.

In addition to the pooled OLS regression analysis, the estimation results of the panel specific models, such as the fixed and random effects models considering AR(1) disturbance term are summarized in Table 3.6. The results are tabulated for each rating scale and the Current Account/GDP and Inflation variables fall into the category of insignificant for both the random and fixed effects models under all estimation analyses with respect to the different rating scales. According to the results of the Hausman specification test, we conclude that the random effects model is the preferable one in our analysis[1]. However, we also put the estimation results of the fixed effect model in Table 3.6 for a robustness check. The explanatory power of the random effects model within countries for each rating scale is nearby 70%. Additionally, the estimated values of parameter ρ for each rating scale are also presented in the table.

To sum up the results of the linear analysis, our findings indicate that the Budget Balance/GDP, GDP per capita, Governance Indicators, and Reserves/GDP variables show up significantly in all linear models and are independent of the rating scale. Significance of Budget Balance/GDP variable can be easily understood by looking at the link between the ratings and budget performances of the countries. Dornbusch (2002) suggests that rating agencies should pay more attention to budget balance due to its role in the experience of past crisis in emerging countries. A positive significant relationship between ratings and GDP per capita of emerging markets is clearly expected because countries, that have more enhanced economies, are anticipated to have more stable debt structures and therefore higher credit ratings. The Effect of Governance Indicators variable, which is the average level of some governance indicators, such as Control of Corruption, Political Stability, and Absence of Violence, can be explained by a pursuit reliability or trust for

[1]The Hausman specification test result changes with respect to the computation method of autocorrelation. For example, we sometimes face with a negative Chi-Square value. However, negative Chi-Square value may be interpreted as preferring to random effects model (Hausman and McFadden (1984)).

the governance structure of emerging markets. Lastly, similar to the findings in Afonso et al. (2007), we indicate that foreign reserves over the GDP variable is one of the most important determinants of the credit rating for emerging markets.

3.4.2 Results of Ordered Response Analysis

Considering the results that we obtained from linear regression in section 3.4.1, we construct the following ordered response model for the analysis of the credit ratings of emerging markets. Table 3.7 gives the estimation results of the standard ordered probit model with clustered robust standard errors. Both the level of inflation for all rating scales and the GDP Growth for Scale 3 are insignificant here. All remaining significant variables have expected signs that we mentioned above. The second model in the ordered response analysis is to use the ordered probit while taking into consideration of random effects. The results with respect to log likelihood maximization using cluster robust standard errors are illustrated in Table 3.8. Here, all significant variables, except the Inflation variable, reflect the signs that we expect. Again, the GDP Growth variable does not appear to be significant for Scale 3. Although the Current Account/GDP and Inflation variables are significant in Scale 1 and 2, they become insignificant in the ordered response analysis of Scale 3.

[Insert Tables 3.7 and 3.8 here]

The ordered response analysis of the credit rating for emerging markets shows that all significant variables in the linear analysis and also External Debt/Export variable have significant effects on rating grades. The negative effect of External Debt per Export on credit ratings points out the impact of affordability of external indebtedness for emerging markets. Additionally, this finding is compatible with various research in the literature (see Cantor and Packer (1996), Bissoondoyal-Bheenick (2005)).

3.5 Post-Estimation Results

To compare the estimation results that we find in the previous section and to investigate the estimation performance of the models that we use in this essay, we focus on the model predictions. First, we deal with predictions of models for all the emerging markets using panel data, after which we make a country specific prediction analysis in the case of Turkey.

3.5.1 Predictions

Due to having the largest variation and one to one transformation, Scale 1 is chosen for evaluation of the prediction made, and the results are given in Table 3.9. Errors in the predictions are tabulated in notches. The category of prediction errors refer to variation of the predicted value from the actual value. Therefore, a positive number in this category means that the predicted value is higher than actual value and vice versa for a negative number.

[Insert Table 3.9 here]

For the linear analysis, a 95% confidence interval is constructed for the fitted value. Afterward, the closest value in this interval to the actual value is selected as the predicted value and then rounded to the nearest integer. However, we use just fitted values for the random effects model with the AR(1) disturbance term due to computational limitations. While the standard pooled OLS model correctly predicts 42% of observations, the random effects model correctly predicts 93% of observations within 4 notches.

For the models in the ordered response analysis, the probability that each observation is in each rating notch is calculated. After calculating the probabilities, the notch which has the maximum predicted probability, is selected as the predicted value. The standard ordered probit model with robust standard errors correctly predicts 36%, but the correct predictions increase to 82% when considering within 4 notches. The performance of the

random effects ordered probit model with robust standard errors is a little better. Those correctly predicted in all observations totals 43% which increases to 93% for all observations when considered within 4 notches.

In order to compare the models in overall, we also add Root Mean Square Error (RMSE) values in Table 3.9 for each model. We reach that there is not much variation between the models. We can compare the models that we use by examining either the values of RMSE or correct predictions within notches. To compare the models comprehensively, RMSE comparison would be preferred, because RMSE searches and scans for all predictions. However, the analysis of correct prediction within notches overlooks the outliers in prediction evaluation, so it can be used to compare models for understanding the correct predictions.

3.5.2 Country specific analysis: Turkey

After looking at the overall performance of the predictions for each model, we want to make a country specific analysis. We use the results of the standard ordered probit model and prediction evaluation for each country indicated in Table 3.10[2]. We compare the prediction errors for the emerging markets employed in this study. According to the results in the table, we obtain that the overall prediction performances of Turkey and India are not as good as the rest of the emerging countries, especially looking within 4 notches. The most interesting point here is the countries' weighted average of prediction errors. The weighted average is constructed by multiplying the prediction error with frequency. The results indicate that the predictions made of Turkey's credit ratings, when compared to other countries, are apparently higher than their actual values. If we look at the prediction errors for Turkey, we see that the model predicts an upgrade of 9 notches for 7 times. That is, S&P has a weak evaluation for Turkey, especially in the highest rating levels. Turkey's

[2]In order to see country specific impacts, we do not prefer to use RE models. Also, ordered response models solves the theoretic caveat in OLS model. Lastly, prediction evaluation for other models give similar results, so here we choose standard ordered probit model.

economic performance in recent years shows that Turkey will be placed in higher rating levels in the following years. Therefore, the results that we find in this study reveal that S&P will have perform poorly in adapting their assessment to Turkey's new era of high ratings.

[Insert Table 3.10 here]

3.6 Concluding Remarks

In recent years, the rating agencies have become more important for various emerging countries since these countries have themselves become more than ready for the financial integration and globalization within the last 10 years. India, China, Brazil, and Turkey, among others, may be thought of in this context. The last financial global crises allowed us the opportunity to consider and question the major rating agencies and their systems. Although we do not explicitly know what weight they give to which variable in their rating assessment, we have discovered a variety of economical, social, and political factors that influence sovereign credit ratings. The latest studies, especially, did this by combining the countries from several categories, such a s developing and developed, high and low ranked, etc. We have, therefore, sufficient evidence to compose either countries or groups of countries although there are not many within the actual country-category or country specific case studies.

In this study, we attempt to verify the validity of these findings for a specific country-category, namely emerging markets. We use the quarterly panel data of credit ratings and key macroeconomic indicators from 2002 to 2011 for 8 emerging markets. Additionally, the credit rating schema of S&P is linearly transformed to three different scales, and it is found that the regressions based on different scales indicate first order autocorrelation in error terms. We deal with first order autocorrelation by using corrected or robust standard errors. In this essay, we separate our analysis into two parts, using linear and ordered

response models. For both models, we present panel specific results using the random effects model. The estimation results of the linear models indicate that the variables of Budget Balance/GDP, GDP per capita, Governance Indicators and Reserves/GDP have a significant effect on credit ratings independent of rating scale. In the ordered response analysis, all significant variables in the linear models plus External Debt/Export variable appear to be significant, and that these results are again independent of the rating scale.

Furthermore, we give post-estimation results, such as predictions of the models for all the emerging markets using panel data. We can correctly predict up to 93% of all observations within a range of 4 notches. We also look the prediction performance of our model in the case of Turkey as compared to other emerging markets. Here, the results indicate that Turkey's actual credit ratings are precisely less than their predicted ratings. S&P's assessment of Turkey's sovereign credit rating exhibits poor performance, especially in the highest rating levels. This leads us to question S&P's timing on upgrading the ratings. Specifically, the following question should be asked: "Do they react fast to bad news and slowly to good news?" Taking all of these facts into account, we believe that S&P's ratings deserve strong criticism.

4 A New Correlation Coefficient for Bivariate Time-Series Data

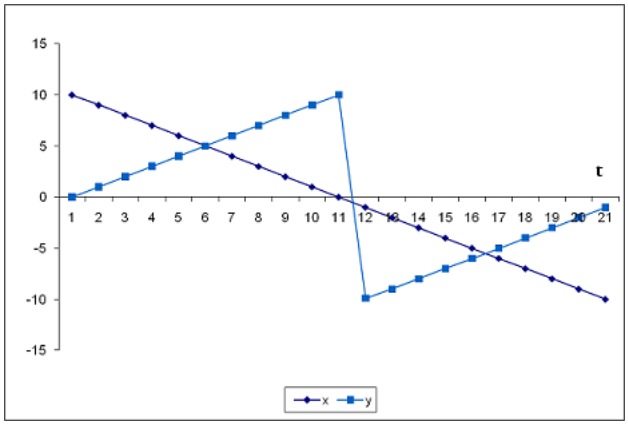
4.1 Introduction

Various financial models (such as pairs trading) are concerned with the correlation between two different time series data, e.g., stock prices or returns. Pearson's product moment correlation coefficient is the most commonly used estimator in measuring such correlations. However, there are many underlying assumptions (such as stationarity) for the validity of this coefficient (Wilcox and Muska (2001)).

If a sample set of time series data is stationary, then the population's mean, variance, and covariance between any two different dates can be estimated based on the sample. If a data is nonstationary, then it violates certain assumptions while estimating these parameters. In general, price series are assumed to be non-stationary, whereas returns are assumed to be stationary. Thus, using Pearson's formula for the calculation of correlation between two price series is not appropriate (Alexander (2008)).

Apart from stationarity, there is another drawback about Pearson's correlation coefficient: It is concerned with the distance of two variables from their means. Assuming that we are interested in two variables that move in the opposite direction while at the same time being both above or below their means. If both of the variables are above (or below) their means, the sum of multiplication of the two variables' deviations from their means will positively contribute to the numerator in Pearson's formula and hence to the correlation coefficient, although the variables move in the opposite direction. The following example (Figure 4.1) may illustrate this problem:

Figure 4.1: An Illustration of Directionality Detection Problem



Note: $\begin{cases} X_t \textit{ points (diamond) are generated as } X_t = 11 - t \textit{ and} \\ Y_t \textit{ points (square) are generated as } Y_t = t - 1 \textit{ for } t = 1 \textit{ to } 11 \textit{ and } Y_t = t - 22 \textit{ for } t = 12 \textit{ to } 21. \end{cases}$

In this example, the means of both of the variables is 0, and they are above their means between $t = 0$ and $t = 11$, below their means between $t = 12$ and $t = 21$. Although, the variables are moving in the opposite direction almost all the time, Pearson's correlation

coefficient, denoted as ρ_p, is 0.50.

A similar idea holds true when two variables move in the same direction while one of the variables is above its mean whereas the other variable is below its mean. In this case, the sum of multiplication of the two variables' deviations from their means will negatively contribute to the numerator in Pearson's formula and hence to the correlation coefficient, even though the variables move in the same direction.

In this article we propose a new correlation coefficient that measures the distance between two subsequent data points by taking the lag difference into consideration. Although the very first data point is lost, we demonstrate that the new correlation coefficient better captures the direction of the covariation of the two variables over time. We also propose various extensions of this coefficient in order to obtain more reasonable and reliable results at the expense of having more complex formulas.

The essay proceeds as follows: We present preliminaries in Section 4.2. In Section 4.3, we introduce the new correlation coefficient, discuss its properties. We exhibit a series of simulations to show the characteristics of the new correlation coefficient in section 4.4. Furthermore, we conclude our work and point to prospective research directions in section 4.5. Finally, we present the definition and some properties of Pearson's correlation coefficient and matrix forms of correlation coefficients in Appendix B.

4.2 Preliminaries

Let $P_{i,t}$ (hereafter P_{it} for today, $P_{i,t-s}$ for a lagged time of s units) represents the price of an asset i at time t. We will denote the entire sequence of values $\{P_{i1}, P_{i2}, ..., P_{iT}\}$ as $\{P_{it}\}$.

The simple return of asset i at time t is defined as:

$$R_{it} = \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}}. \tag{38}$$

Similarly log-return is defined as:

$$r_{it} = \log\left(P_{it}/P_{i,t-1}\right). \tag{39}$$

Let $\{X_t\}$ be a kind of stochastic process, we define the stationarity as follows: A stochastic process $\{X_t\}$ having a finite mean and variance is said to be stationary, if for all t and $t-s$:

$$\mathbf{E}\left(X_t\right) = \mathbf{E}\left(X_{t-s}\right) = \mu \tag{40}$$

$$\mathbf{E}\left[\left(X_t-\mu\right)^2\right] = \mathbf{E}\left[\left(X_{t-s}-\mu\right)^2\right] = \sigma^2 \tag{41}$$

$$\mathbf{E}\left[\left(X_t-\mu\right)\left(X_{t-s}-\mu\right)\right] = \mathbf{E}\left[\left(X_{t-j}-\mu\right)\left(X_{t-j-s}-\mu\right)\right] = \gamma_{(s)} \tag{42}$$

where $\mu, \sigma^2, \gamma_{(s)}$ are all constants i.e., independent of time (Enders (2004)).

In practice, stock prices may be assumed as non-stationary, whereas simple and log returns may be assumed to be stationary (Alexander (2008)). Furthermore, conventionally logarithm of stock prices are assumed to follow *Geometric Brownian Motion* (Tsay (2005)) which means that log-returns are assumed to be normally distributed.

4.3 The New Correlation Coefficient and Its Properties

Since Pearson's correlation coefficient, ρ_P, might have some problems for time-series data, we suggest the following correlation coefficient as an alternative:

$$\rho_O = \frac{\alpha_{xy}}{\alpha_x \alpha_y} \tag{43}$$

where $\alpha_x^2 = \mathbf{E}\left[\left(X_t - X_{t-1}\right)^2\right]$, $\alpha_y^2 = \mathbf{E}\left[\left(Y_t - Y_{t-1}\right)^2\right]$ and $\alpha_{xy} = \mathbf{E}\left[\left(X_t - X_{t-1}\right)\left(Y_t - Y_{t-1}\right)\right]$.

The new correlation coefficient ρ_O can then be estimated as:

$$\widehat{\rho}_O = \frac{A_{xy}}{A_x A_y} \tag{44}$$

where $A_x^2 = \frac{1}{(T-1)}\sum_{t=2}^{T}(X_t - X_{t-1})^2$, $A_y^2 = \frac{1}{(T-1)}\sum_{t=2}^{T}(Y_t - Y_{t-1})^2$ and,

$A_{xy} = \frac{1}{(T-1)}\sum_{t=2}^{T}(X_t - X_{t-1})(Y_t - Y_{t-1})$ which are related to the first order autocorrelations of X_t and Y_t, respectively (Brockwell and Davis (2009)). It should also be noticed that α_{xy} is estimated by A_{xy} and α_x, α_y by A_x, A_y, respectively. For any two processes $\{X_t\}$ and $\{Y_t\}$, it is easy to see that $\mathbf{E}(A_x^2) = \alpha_x^2$, $\mathbf{E}(A_y^2) = \alpha_y^2$, and $\mathbf{E}(A_{xy}) = \alpha_{xy}$.

If we let $X_t = \log(P_{1t})$ for the first stock price and $Y_t = \log(P_{2t})$ for the second stock price, then Equation (44) can be written as:

$$\widehat{\rho}_O = \frac{\sum_{t=2}^{T} r_{1t} r_{2t}}{\sqrt{\sum_{t=2}^{T} r_{1t}^2 \sum_{t=2}^{T} r_{2t}^2}} \tag{45}$$

where r_{1t}, r_{2t} are defined in Equation (39).

4.3.1 Some Properties of the New Correlation Coefficient

It can be shown that ρ_O satisfies the inequality $|\rho_O| \leq 1$. Using the Cauchy-Schwartz inequality, one can write that the random variables (processes) X_t, Y_t satisfy:

$$\{\mathbf{E}[(X_t - X_{t-1})(Y_t - Y_{t-1})]\}^2 \leq \mathbf{E}\left[(X_t - X_{t-1})^2\right]\mathbf{E}\left[(Y_t - Y_{t-1})^2\right] \tag{46}$$

with equality holding if and only if one of the variables is almost surely a multiple of the other, that is, $P[a(X_t - X_{t-1}) = b(Y_t - Y_{t-1})] = 1$ for some real a and b, at least one of which is non-zero.

We can rewrite Equation (46) as:

$$-1 \leq \frac{\mathbf{E}\left[\left(X_t - X_{t-1}\right)\left(Y_t - Y_{t-1}\right)\right]}{\sqrt{\mathbf{E}\left[\left(X_t - X_{t-1}\right)^2\right]\mathbf{E}\left[\left(Y_t - Y_{t-1}\right)^2\right]}} = \rho_O \leq 1 \tag{47}$$

since $\mathbf{E}\left[\left(X_t - X_{t-1}\right)^2\right]$ and $\mathbf{E}\left[\left(Y_t - Y_{t-1}\right)^2\right]$ are strictly positive.

4.3.2 The Properties Under Stationarity

Next, we derive the quantities α_x^2, α_y^2, and α_{xy} under the bivariate stationarity assumption for $\{X_t,Y_t\}$, that is, under the assumption that:

$$\mathbf{E}\left(X_t\right) = \mathbf{E}\left(X_{t-s}\right) = \mu_X, \tag{i}$$

$$\mathbf{E}\left(Y_t\right) = \mathbf{E}\left(Y_{t-s}\right) = \mu_Y,$$

$$\mathbf{E}\left[\left(X_t - \mu_X\right)^2\right] = \mathbf{E}\left[\left(X_{t-s} - \mu_X\right)^2\right] = \sigma_X^2, \tag{ii}$$

$$\mathbf{E}\left[\left(Y_t - \mu_Y\right)^2\right] = \mathbf{E}\left[\left(Y_{t-s} - \mu_Y\right)^2\right] = \sigma_Y^2,$$

$$\mathbf{E}\left[\left(X_t - \mu_X\right)\left(X_{t-s} - \mu_X\right)\right] = \gamma_X(s), \tag{iii}$$

$$\mathbf{E}\left[\left(Y_t - \mu_Y\right)\left(Y_{t-s} - \mu_Y\right)\right] = \gamma_Y(s).$$

Furthermore, the cross-covariance between $\{X_t\}$ and $\{Y_t\}$ at lag s (Brockwell and Davis (2009)) is;

$$\mathbf{E}\left[\left(X_t - \mu_X\right)\left(Y_{t-s} - \mu_Y\right)\right] = \gamma_{XY}(s). \tag{iv}$$

Notice that at lag 0 (i.e., $s = 0$), $\gamma_X(0) = \sigma_X^2$, $\gamma_Y(0) = \sigma_Y^2$, and $\gamma_{XY}(0) = \sigma_{XY}^2 = \mathbf{Cov}\left(X, Y\right)$.

Then,

$$\begin{aligned}\alpha_x^2 &= \mathbf{E}\left[(X_t - X_{t-1})^2\right] = \mathbf{E}\left[X_t^2 - 2X_tX_{t-1} + X_{t-1}^2\right] \\ &= \mathbf{E}\left[X_t^2\right] - 2\mathbf{E}\left[X_tX_{t-1}\right] + \mathbf{E}\left[X_{t-1}^2\right] \\ &= 2\mathbf{E}\left[X_t^2\right] - 2\mathbf{E}\left[X_tX_{t-1}\right] \\ &= 2\left(\sigma_X^2 + \mu_X^2\right) - 2(\gamma_X(1) + \mu_X^2) \\ &= 2\left(\sigma_X^2 - \gamma_X(1)\right)\end{aligned} \tag{48}$$

since $\mathbf{E}\left[X_t^2\right] = (\sigma_X^2 + \mu_X^2)$ and $\mathbf{E}\left[X_tX_{t-1}\right] = \gamma_X(1) + \mu_X^2$. Similarly, $\alpha_y^2 = 2\left(\sigma_Y^2 - \gamma_Y(1)\right)$.

Furthermore,

$$\begin{aligned}\alpha_{xy} &= \mathbf{E}\left[(X_t - X_{t-1})\left(Y_t - Y_{t-1}\right)\right] \\ &= \mathbf{E}\left[X_tY_t\right] - \mathbf{E}\left[X_tY_{t-1}\right] - \mathbf{E}\left[X_{t-1}Y_t\right] + \mathbf{E}[X_{t-1}Y_{t-1}] \\ &= 2\mathbf{E}[X_tY_t] - 2\mathbf{E}[X_tY_{t-1}] \\ &= 2\left(\gamma_{XY}(0) + \mu_X\mu_Y\right) - 2(\gamma_{XY}(1) + \mu_X\mu_Y) \\ &= 2\left(\gamma_{XY}(0) - \gamma_{XY}(1)\right)\end{aligned} \tag{49}$$

since $\mathbf{E}[X_tY_t] = \gamma_{XY}(0) + \mu_X\mu_Y$ and $\mathbf{E}[X_tY_{t-1}] = \gamma_{XY}(1) + \mu_X\mu_Y$.

But, since $\gamma_{XY}(0) = \sigma_{XY}^2$ and $\sigma_{XY}^2 = \rho_{XY}\sigma_X\sigma_Y$ where $\rho_{XY} = \mathbf{Corr}(X, Y)$ which equals ρ_P. Then,

$$\alpha_{xy} = 2(\rho_P\sigma_X\sigma_Y - \gamma_{XY}(1))$$

Thus, for stationary bivariate time series $\{X_t, Y_t\}$,

$$\rho_O = \frac{\rho_P\sigma_X\sigma_Y - \gamma_{XY}(1)}{\sqrt{\left(\sigma_X^2 + \gamma_X(1)\right)\left(\sigma_Y^2 + \gamma_Y(1)\right)}} \tag{50}$$

Observe that when $\gamma_X(1) = \gamma_Y(1) = \gamma_{XY}(1) = 0$,

$$\rho_O = \frac{\rho_P \sigma_X \sigma_Y}{\sigma_X \sigma_Y} = \rho_P$$

Hence, ρ_O and ρ_P agree whenever autocorrelations[1] of $\{X_t\}$ and $\{Y_t\}$ at lag 1 and cross-correlation of $\{X_t\}$ and $\{Y_t\}$ at lag 1 are all zero. Furthermore, under the bivariate stationarity, ρ_P ignores the autocorrelations and cross-correlation, whereas ρ_O incorporates lag 1 autocorrelations.

If $\{X_t\}$ and $\{Y_t\}$ are each (univariate) stationary as in section 4.2, but $\{X_t\}$ and $\{Y_t\}$ are independent that is, above (i)-(iii) hold and (iv) holds with $\gamma_{XY}(s) = 0$ for all $s \geq 0$, then $\rho_P = 0$, hence

$$\rho_O = \frac{\rho_P \sigma_X \sigma_Y}{\sqrt{(\sigma_X^2 + \gamma_X(1))(\sigma_Y^2 + \gamma_Y(1))}} = 0 \tag{51}$$

If X_t are i.i.d. with $\mathbf{E}(X_t) = \mu_X$, $\mathbf{Var}(X_t) = \sigma_X^2$ and Y_t are i.i.d. with $\mathbf{E}(Y_t) = \mu_Y$, $\mathbf{Var}(Y_t) = \sigma_Y^2$, then

$$\rho_O = \frac{\rho_P \sigma_X \sigma_Y - \gamma_{XY}(1)}{\sigma_X \sigma_Y} = \rho_P - \frac{\gamma_{XY}(1)}{\sigma_X \sigma_Y} \tag{52}$$

since $\gamma_X(1) = \gamma_Y(1) = 0$ for i.i.d $\{X_t\}$ and $\{Y_t\}$.

If $\{X_t\}$ and $\{Y_t\}$ are both i.i.d. as above and are independent of each other then $\rho_O = \rho_P = 0$, since $\gamma_{XY}(1) = 0$ for independent $\{X_t\}$ and $\{Y_t\}$.

Therefore, we once again state that ρ_O and ρ_P measure related but different things. Specifically, they both measure the covariation of $\{X_t\}$ and $\{Y_t\}$ but with emphasis on different aspects. For example, in a hypothesis testing framework, we have $\rho_O = \rho_P = 0$ under independence of $\{X_t\}$ and $\{Y_t\}$. However, if $\{X_t\}$ and $\{Y_t\}$ are independent only at lag 0, but dependent at lag 1 (i.e., $\gamma_{XY}(s) \neq 0$ for $s = 1$), then $\rho_P = 0$ but $\rho_O = \frac{-\gamma_{XY}(1)}{\sigma_X \sigma_Y}$.

[1]Here, the term of autocorrelation is used based on the definition of Pearson's correlation coefficient. Since first lagged variable is in the formula of the new correlation coefficient, we do not prefer to apply the definition of the new correlation coefficient to measure autocorrelation especially at lag 1. For any *i.i.d.* X_t, autocorrelation of X_t at lag $n \geq 2$ with respect to new correlation coefficient is $\rho_O(n) = 0$.

4.3.3 The Properties Under Non-Stationarity

In this section, we consider special cases under the bivariate non-stationarity assumption for $\{X_t,Y_t\}$. We investigate the properties of the new correlation coefficient in two special cases: namely, Spurious Correlation and Cointegration. Then, we simulate both the new and Pearson's correlation coefficients for these cases in section 4.4.

Spurious Correlation: In this case, X_t and Y_t are generated by the independent random walks:

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t^x, \\ Y_t &= Y_{t-1} + \varepsilon_t^y, \end{aligned} \tag{53}$$

in which ε_t^x are $i.i.d.(0, \sigma_{\varepsilon^x}^2)$ and ε_t^y are $i.i.d.(0, \sigma_{\varepsilon^y}^2)$. In our simulations in Section 4.4, we draw ε_t^x and ε_t^y from independent $N(0,1)$ populations. As T goes to infinity, the numerator of the new correlation coefficient in (6) goes to zero (Phillips (1986)).

$$\alpha_{xy} = \mathbf{E}\left[(X_t - X_{t-1})\,(Y_t - Y_{t-1})\right] = \mathbf{E}\left[(\varepsilon_t^x)\,(\varepsilon_t^y)\right] \to 0. \tag{54}$$

Therefore, the new correlation coefficient also goes to zero.

$$\rho_O = \frac{\alpha_{xy}}{\alpha_x \alpha_y} \to 0 \qquad a.s. \tag{55}$$

Cointegration: Now we have two cointegrated non-stationary variables which are generated by,

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t^x, \\ Y_t &= \alpha X_t + \varepsilon_t^y, \end{aligned} \tag{56}$$

in which ε_t^x are $i.i.d.(0, \sigma_{\varepsilon^x}^2)$ and ε_t^y are $i.i.d.(0, \sigma_{\varepsilon^y}^2)$. As T goes to infinity

$$\rho_O = \frac{\mathbf{E}\left[(X_t - X_{t-1})(Y_t - Y_{t-1})\right]}{\sqrt{\mathbf{E}\left[(X_t - X_{t-1})^2\right]\mathbf{E}\left[(Y_t - Y_{t-1})^2\right]}} \to \frac{1}{\sqrt{1 + \frac{2}{\alpha^2}\frac{\sigma_{\varepsilon^y}^2}{\sigma_{\varepsilon^x}^2}}} \tag{57}$$

since,

$\mathbf{E}\left[(\varepsilon_t^x)^2\right] = \sigma_{\varepsilon^x}^2$ as $T \to \infty$,

$\mathbf{E}\left[(\varepsilon_t^y)^2\right] = \mathbf{E}\left[\left(\varepsilon_{t-1}^y\right)^2\right] = \sigma_{\varepsilon^y}^2$ as $T \to \infty$.

4.3.4 The Properties in General Context

Here, we introduce two propositions and show their proofs in general circumstances. The state of the independent variables and the essential properties of the new correlation coefficient enable us to make two propositions.

Proposition 1. *If X_t, Y_t are one lag cross-independent[2] and also if at least one of X_t and Y_t is not divergent[3], then the new correlation coefficient $\rho_O = 0$.*

Proof. The numerator of ρ_O is α_{xy} and equals to:

$$\alpha_{xy} = \mathbf{E}\left[(X_t - X_{t-1})(Y_t - Y_{t-1})\right] \tag{58}$$

Because of the linearity property of the expected value, Equation (58) becomes as follows:

$$\alpha_{xy} = \mathbf{E}\left[X_t Y_t\right] - \mathbf{E}\left[X_t Y_{t-1}\right] - \mathbf{E}\left[X_{t-1} Y_t\right] + \mathbf{E}\left[X_{t-1} Y_{t-1}\right] \tag{59}$$

[2]We define one lag cross independence between two variables X_t and Y_t, if pairs of $\{X_{t-m}, Y_{t-n}\}$ for $m, n = 0, 1$; are pairwise independent.

[3]Here, divergence of a variable Z_t is defined when $\mathbf{E}(Z_t) \neq \mathbf{E}(Z_{t-1})$.

Since X_t and Y_t are one lag cross-independent, the RHS of Equation (59) turns into:

$$\alpha_{xy} = \mathbf{E}[X_t]\,\mathbf{E}[Y_t] - \mathbf{E}[X_t]\,\mathbf{E}[Y_{t-1}] - \mathbf{E}[X_{t-1}]\,\mathbf{E}[Y_t] + \mathbf{E}[X_{t-1}]\,\mathbf{E}[Y_{t-1}] \tag{60}$$

By the non-divergence of X_t, $\mathbf{E}(X_t) = \mathbf{E}(X_{t-1})$. Equation (60) can be rewritten as:

$$\alpha_{xy} = \mathbf{E}[X_t]\,\mathbf{E}[Y_t] - \mathbf{E}[X_t]\,\mathbf{E}[Y_{t-1}] - \mathbf{E}[X_t]\,\mathbf{E}[Y_t] + \mathbf{E}[X_t]\,\mathbf{E}[Y_{t-1}] \tag{61}$$

$$= 0 \tag{62}$$

Therefore, $\alpha_{xy} = 0$, implying $\rho_O = 0$ □

This proposition tells us that whenever there is at least one non-divergent variable, one lag cross-independence implies a zero new correlation coefficient. However, Pearson's correlation coefficient may not capture the cross-independence. As such, it may not be zero when at least one of the variables is not stationary. The condition of non-divergence is stronger than the stationarity condition. Not only stationary variables, but also some non-stationary variables such as random walks are non-divergent. Therefore, the new correlation coefficient is capable of observing the one lag cross-independence between the variables, even if the variables are non-stationary. Furthermore, the contra-positive of this proposition is also beneficial. If the new correlation coefficient is not zero, this refers to a violation of the one lag cross-independence condition or two divergent variables which are non-stationary.

Proposition 2. *If $\rho_O = 0$, then X_t, Y_t are not one lag cross-independent or at least one of X_t and Y_t is not divergent.*

Proof. Suppose for a contradiction that the imposed condition does not hold. So we have that X_t, Y_t are one lag cross-independent,and none of X_t and Y_t is not divergent, that is both of X_t and Y_t are divergent.

By the derivation of the previous proposition, the one lag cross-independence indicates that:

$$\alpha_{xy} = \mathbf{E}[X_t]\,\mathbf{E}[Y_t] - \mathbf{E}[X_t]\,\mathbf{E}[Y_{t-1}] - \mathbf{E}[X_{t-1}]\,\mathbf{E}[Y_t] + \mathbf{E}[X_{t-1}]\,\mathbf{E}[Y_{t-1}] \tag{63}$$

Furthermore, divergent X_t and Y_t mean that:

$$\mathbf{E}(X_t) = \mathbf{E}(X_{t-1}) + \alpha \tag{64}$$

$$\mathbf{E}(Y_t) = \mathbf{E}(Y_{t-1}) + \beta \tag{65}$$

where α and β are any real number and not equal to zero.

Therefore, Equation (63) may be rewritten as:

$$\begin{aligned} \alpha_{xy} &= [\mathbf{E}[X_{t-1}] + \alpha][\mathbf{E}[Y_{t-1}] + \beta] - [\mathbf{E}[X_{t-1}] + \alpha]\mathbf{E}[Y_{t-1}] - ... \\ &\quad ... \ \mathbf{E}[X_{t-1}]\,[\mathbf{E}[Y_{t-1}] + \beta] + \mathbf{E}[X_{t-1}]\,\mathbf{E}[Y_{t-1}] \qquad (66) \\ &= \alpha\beta \qquad (67) \end{aligned}$$

That is to say that the numerator of ρ_O is not zero ($\alpha_{xy} = \alpha\beta \neq 0$). Also, the denominator of ρ_O is positive, so ρ_O is not zero which contradicts our imposing statement that says that $\rho_O = 0$. □

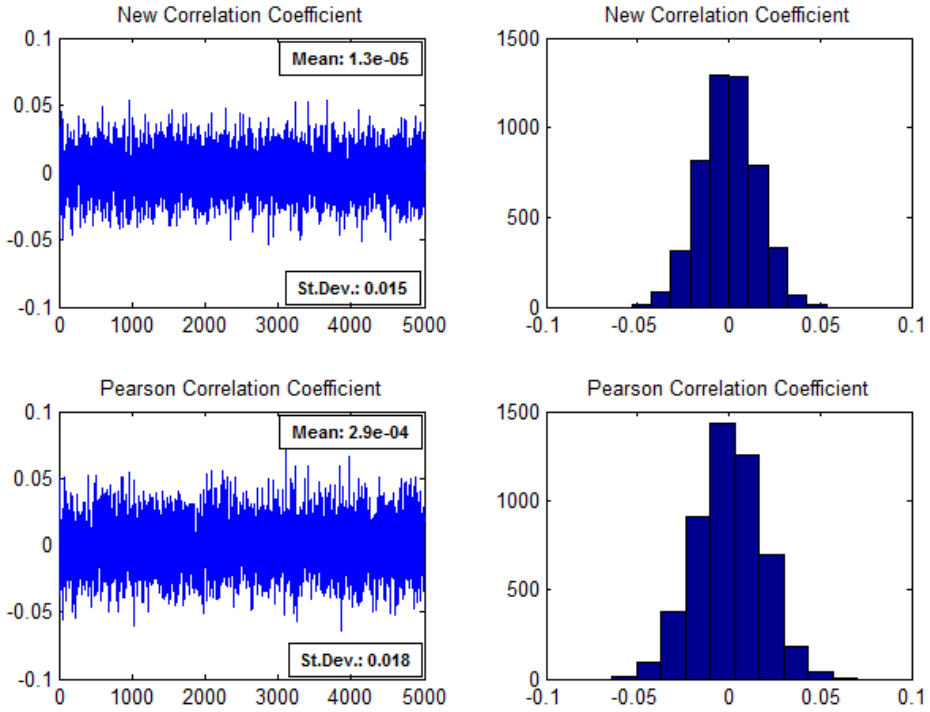
Here, this proposition is similar to the converse of the previous proposition. We try to figure out what a zero new correlation coefficient implies. We find that if the new correlation coefficient is zero, then X_t and Y_t do not provide the condition of one lag cross-independence or at least one of the variables is not divergent. Therefore, a zero new correlation coefficient indicates that the one lag cross-independence condition is violated or that there is a minimum of one non-divergent variable.

4.4 Monte Carlo Simulations

4.4.1 Stationary Case

Here we report a Monte Carlo simulation to compare the new correlation coefficient and Pearson's correlation coefficient for one of the cases of the stationary variables which is stated in Section 4.3.2. In our Monte Carlo simulation setup, we form 5000 simulations and for each simulation 5000 data points are generated for both of the variables. We employ the case of two stationary variables which are both i.i.d. and independent of each other. The results of the simulation are given in Table 4.2.

Figure 4.2: Plots and Histograms of Correlation Coefficients in the Case of Independent Stationary Variables



Note: The plot in the top left and the histogram in the top right show how the new correlation coefficient is distributed. The plot and histogram in the bottom are for the Pearson's correlation coefficient.

As we mention in Section 4.3.2, both correlation coefficients go to zero because these vari-

ables are two i.i.d. stationary variables and independent of each other. When we compare the two correlation coefficients, we reach that the mean value of the new correlation coefficient is closer to zero than Pearson's correlation coefficient. So the new one is faster in reaching to zero compared to Pearson's one. Additionally, the standard deviation of the new correlation coefficient is less than the Pearson's correlation coefficient. As a result, the new correlation coefficient dominates the Pearson's correlation coefficient in terms of capturing the independence among the variables.

4.4.2 Non-Stationary Cases

In this section, we perform Monte Carlo simulations to illustrate some of the properties of the new correlation coefficient compared to the Pearson's correlation coefficient for various non-stationarity cases.

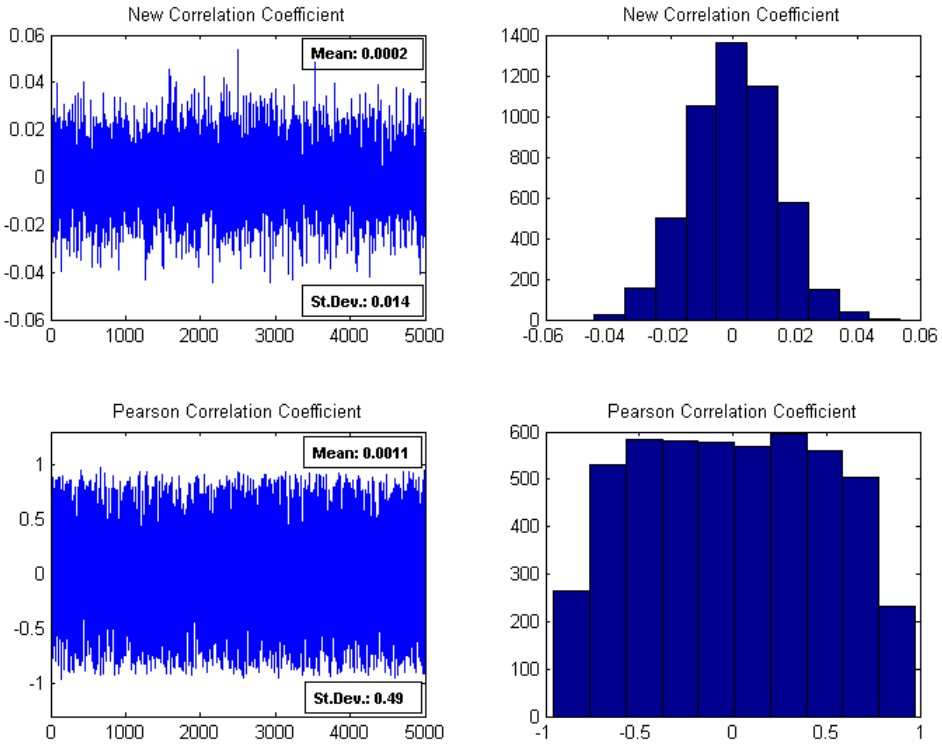
Two Random Walks (Spuriously Correlated) Here, we take two nonstationary variables, such as two random walks:

$X_t = X_{t-1} + \varepsilon_t^x$,where $\varepsilon_t^x \sim N(0,1)$

$Y_t = Y_{t-1} + \varepsilon_t^y$,where $\varepsilon_t^y \sim N(0,1)$

which can be spuriously correlated or not cointegrated, also the distributions ε_t^x and ε_t^y are independent. Again, we form 5000 simulations and for each simulation 5000 data points are generated for both of the variables X_t and Y_t where X_0 and Y_0 are given as arbitrary numbers. The results of the simulations are shown in Figure 4.3. The differences between the new and Pearson's correlation coefficients in the case of spurious correlation can be easily seen by looking at the plots and histograms of the coefficients.

Figure 4.3: Plots and Histograms of Correlation Coefficients in the Case of Spurious Correlation



Note: The plot in the top left and the histogram in the top right show how the new correlation coefficient is distributed. The plot and histogram in the bottom are for the Pearson's correlation coefficient.

In this case, it is expected that the correlation of the two nonstationary variables should be around zero because these variables are two independent random walks. The Pearson's correlation coefficient on the average is zero, but has extremely large fluctuations (i.e., it could imply a very strong positive or negative correlation for some of the simulated data sets). This issue is called as spurious correlation in the literature (Engle and Granger (1987)). However, the new correlation coefficient takes into account the nonstationarities of the variables, so it provides better results. The values of the simulations of the new correlation coefficient are more clustered around zero and have a shape more similar to normal distribution. Additionally, when we take two random walks with drift, we obtain similar results (not presented).

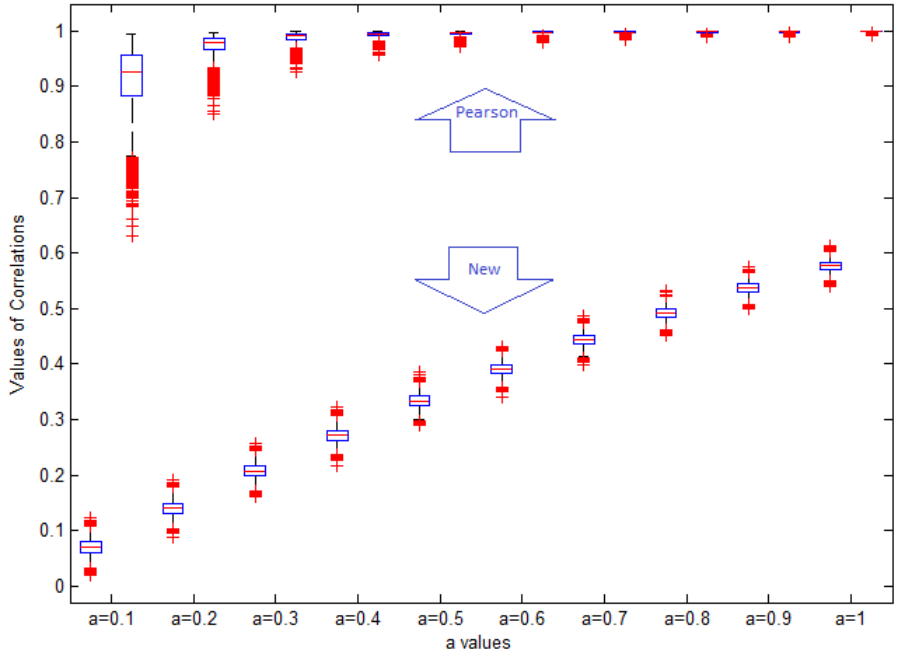
Two Cointegrated Variables We take two nonstationary variables as:

$X_t = X_{t-1} + \varepsilon_t^x$,where $\varepsilon_t^x \sim N(0,1)$

$Y_t = aX_t + \varepsilon_t^y$,where $\varepsilon_t^y \sim N(0,1)$

in which $\{X_t\}$ and $\{Y_t\}$ are cointegrated. For $a = 0.1, 0.2, 0.3, ..., 1$, we use the same Monte Carlo simulation setup as in the previous case (in section 4.4.1). Figure 4.4 shows the box-plot of 5000 simulations for both the new and Pearson's correlation coefficients at each value of a. The value of the Pearson's correlation coefficient is significantly higher than the value of the new correlation coefficient for each value of a. These differences in Figure 4.4 can be explained by the estimations of correlations and the parameters in these estimations. Actually, both of correlations capture the cointegration but give different values according to the parameters and a's. For example, the new correlation coefficient is calculated according to the Equation (57) above, and it is 0.58 when a=1.

Figure 4.4: Box-plots of Correlation Coefficients in the Case of Cointegration



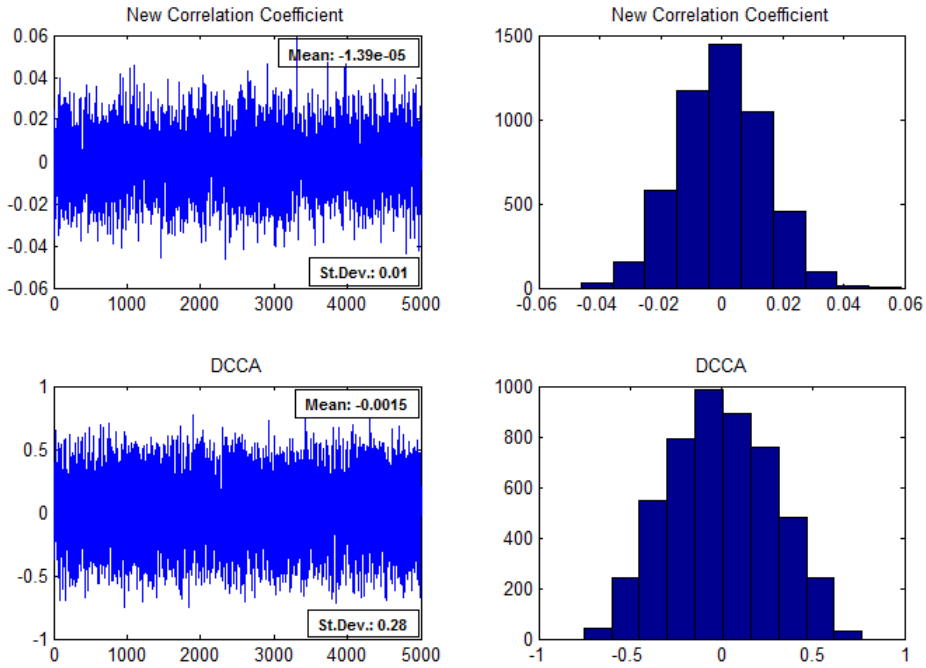
Note: Box-plots represent the new and Pearson's correlation coefficients for different a values.

4.4.3 Comparison of New Correlation Coefficient and DCCA

Correlations based on cross-covariance analysis between bivariate time series has become a trend in Finance and Economics over recent years. The main intuition behind the construction of these correlations is to separation of time series into overlapping boxes of length (scale). Then the correlation analysis is composed of aggregation of the information coming from separated time scales. One of the such correlations is called detrended cross-correlation (DCCA). For non-stationary time series, it is shown that the DCCA dominates the Pearson's correlation coefficient (see Podobnik and Stanley (2008), Zebende (2011) and Kristoufek (2013)). Here we also compare our results with the DCCA values for the two special cases of non-stationary variables.

First, we again take two independent nonstationary variables which are illustrated by two random walks. Similar to the simulation setup in the previous subsection, we perform 5000 simulations. For each simulation, 5000 data points are generated for both of the variables. The length of the overlapping boxes calculating the DCCA is chosen as the half of data points, which is 2500. The results of the simulations are shown in Figure 4.5. When we compare the results for the new correlation with DCCA, we reach that both correlation coefficients capture zero correlation in the case of two independent random walks. Moreover, the standard deviation of the DCCA is found as 0.28. The results in Kristoufek (2013) also indicate that if there are two independent random walks, the standard deviation of the DCCA can take values from 0.08 to 0.40 with respect to changing length of boxes. However, the new correlation coefficient has standard deviation of 0.014 and five times less than the smallest standard deviation of DCCA. Therefore, the new correlation coefficient dominates the DCCA in the case of two independent random walks.

Figure 4.5: Comparison of New and DCCA Correlation Coefficients in the Case of Spurious Correlation



Note: The plot in the top left and the histogram in the top right show how the new correlation coefficient is distributed. The plot and histogram in the bottom are for the Pearson's correlation coefficient.

Additionally, we generate two cointegrated variables with the same setting described in the previous subsection. Both of the correlation coefficients again capture the relationship between the variables depending the parameters selection. Kristoufek (2013) illustrates that the standard deviation of the DCCA is in between 0.01 and 0.1 with respect to changing length of boxes. For the same case, the standard deviation of the new correlation coefficient is 0.01 which is the same as the smallest standard deviation of the DCCA. That is, the new correlation coefficient has almost similar performance with the DCCA in the case of cointegration. As a consequence, the new correlation coefficient captures the cases above better than the DCCA correlation coefficient in total.

4.5 Concluding Remarks

Correlation in social sciences, especially in finance, has gained considerable attention recently. Several financial instruments and methods such as pair trading, credit risk applications etc. focus on the correlation between various pairs of variables within data sets. One of the measures of inference for the correlation is the Pearson's correlation coefficient, which requires the following assumptions; normality, linearity, randomness, stationarity and homoscedasticity. Unfortunately, these assumptions are rarely satisfied for real data. Pearson's correlation coefficient is not appropriate, if assumptions such as stationarity do not hold. The pitfalls of the Pearson's correlation coefficient motivated us to introduce a new correlation coefficient which is measured by taking into account the lag difference of data points. We demonstrate that the new correlation coefficient has better performance in capturing the cross-independence of two variables over time. Additionally, we illustrated the key differences between Pearson's and the new correlation coefficients under a number of cases with Monte Carlo simulations. The most notable improvements of the new correlation coefficient can be seen in the situation of the two nonstationary variables which are spuriously correlated. Our simulation study suggests that the new correlation coefficient is more similar to normal distribution as compared to the Pearson's correlation coefficient in the cases we consider. As a robustness check, we also compare the new correlation coefficient with the DCCA for the two cases of the non-stationary variables, and find that the new one dominates the DCCA. For future work, our main goal is to find the (asymptotic) distribution of the new correlation coefficient. Moreover, new simulations for various cases will be conducted, and their results will also be compared.

5 Conclusion

Studies in the area of Macroeconometrics bring various implications for different questions. These implications can be provided by using empirical works, theoretical studies, or both.

This thesis contributes to the literature by providing two empirical and one theoretical essays. With this research, we try to find answers to their respective questions and construct a scientific merit.

The first essay aims to investigate the housing market by looking at both demand and supply sides. We deal with optimization problem of a representative agent in both sides. We also strengthen our model by adding an empirical analysis. The cointegration analysis and the vector error correction model (VECM) are used to identify the impacts of the fundamental macroeconomic factors on the Turkish housing market between October 2007 and June 2012. We find that the housing market and macroeconomic fundamentals, such as interest rate, gross domestic product (GDP), housing prices, among others are cointegrated. On the demand side, we demonstrate that the mortgage credit volume is affected by real house prices, real GDP per capita, expected inflation, and nominal interest rate. On the other hand, it is found on the supply side that the construction permit has a long-run relation with real house prices, real construction cost, and real interest rate. Additionally, real mortgage credit volume and construction permit have adjustments on the disequilibrium of housing demand and that construction permit adjusts the disequilibrium of the housing supply. Furthermore, the impulse response analysis using a structural VECM suggests that the housing market in Turkey is sensitive to shocks in the economy. The first essay also presents a forecast error variance decompositions (FEVD), showing how both real house prices and the real GDP per capita on the housing market are important in the long run.

In the second essay, we examine the factors behind the sovereign ratings of the emerging markets assigned by Standard and Poor's. Although we do not explicitly know the rating assessment of S&P, we discover a variety of economical, social, and political factors influence the sovereign credit ratings of emerging markets. Since the rating agencies' performances about the sovereign ratings of emerging markets bring up some objections, we wanted to verify the sovereign credit ratings within the category of emerging markets.

The data used in this essay is collected on a quarterly base consisting of credit ratings and key macroeconomic indicators from 2002 to 2011 for 8 emerging markets. We employ both a linear and ordered response model in order to conduct an empirical analysis. Our results indicate that the Budget Balance/GDP, GDP per capita, Governance Indicators, and Reserves/GDP are the most relevant factors in explaining credit ratings. Furthermore, we correctly predicted up to 93% of all credit rating levels. The Last but interesting result shows that the prediction performance of our model becomes poor in the case of Turkey as compared to other emerging markets, especially in the highest rating levels.

The last essay introduces a new correlation coefficient for the bivariate time series data. Real data in the Social Sciences does not always satisfy the main assumptions of Pearson's correlation coefficient. For example, Pearson's correlation coefficient is not appropriate if the stationarity of the data does not hold. To avoid this weakness of Pearson's correlation coefficient, we propose a new correlation coefficient in this study. It takes into account the lag difference of two variables instead of the distance of variables from their respective means. It is found that the new correlation coefficient performs better in terms of showing the direction of the covariation of the bivariate time series data, especially when they are independent. Additionally, we investigate some cases using Monte Carlo simulations. In the case of two non-stationary variables which are spuriously correlated, using the new correlation coefficient manifests improved results. Our simulations also suggest that the new correlation coefficient is more similar to the normal distribution as compared to Pearson's correlation coefficient for the special cases that we have dealt with in this essay. Also, we compare the performances of the new correlation coefficient and DCCA for the cases of the non-stationary variables, and reach that the new one dominates the DCCA. When we consider our work and the results we find in here, we believe that the new correlation coefficient will gain a considerable amount of attention in the future.

6 References

1. Abelson P., Joveux R., Milunovich G., and D. Chung. 2005. Explaining House Prices in Australia: 1970-2003. *Economic Record*, 81, 96-103.

2. Afonso A. 2003. Understanding the Determinants of Sovereign Debt Ratings: Evidence for the Two Leading Agencies. *Journal of Economics and Finance,* 27, 56-74.

3. Afonso A., Gomez P., and P. Rother. 2007. What Hides Behind Sovereign Ratings? *European Central Bank.* Working Paper No: 711.

4. Alexander C. 2008. *Quantitative Methods in Finance.* John Wiley.

5. Alper K., Mutluer D., Karasahin R., and H. Atasoy. 2011. Arzın Merkezine Seyahat: Bankacılarla Yapılan Görüşmelerden Elde Edilen Bilgilerle Türk Bankacılık Sektörünün Davranışı. *Central Bank of the Republic of Turkey.*

6. Baer W.C. 1986. The dynamics of the nation's housing stock. *Scientific American,* 255, 29-35.

7. Bickel P.J., and A.K. Doksum. 1977. *Mathematical Statistics, Basic Ideas and Selected Topics.* Prentice Hall.

8. Bissoondoyal-Bheenick E. 2005. An Analysis of the Determinants of Sovereign Ratings. *Global Finance Journal*, 15, 251-80.

9. Bissoondoyal-Bheenick E., Brooks R., and A.Y.N. Yip. 2006. Determinants of Sovereign Ratings: A Comparison of Case-Based Reasoning and Ordered Probit Approaches. *Global Finance Journal*, 17, 136-154.

10. Brissimis S.N., and T. Vlassopoulos. 2009. The Interaction between Mortgage Financing and Housing Prices in Greece. *The Journal of Real Estate Finance and Economics*, 39, 146-164.

11. Brockwell P.J., and R.A. Davis. 2009. *Time Series: Theory and Methods.* Springer.

12. Bulut Z.B. 2009. Demand and Supply of Real Estate Market in Turkey: A Cointegration Analysis. Master Thesis. Bilkent University: Turkey.

13. Cantor R., and F. Packer. 1996. Determinants and Impact of Sovereign Credit Ratings. FRBNY Economic Policy Review, 1, 37-53.

14. Carliner G. 1973. Income Elasticity of Housing Demand. *Review of Economics and Statist*ics, 55, 528-532.

15. Chen M.C., and K. Patel. 1998. House Price Dynamics and Granger Causality: An Analysis of Taipei New Dwelling Market. *Journal of the Asian Real Estate Society*, 1, 101-126.

16. De Leeuw F. 1971. The Demand for Housing: A Review of Cross-Section Evidence. *Review of Economics and Statist*ics, 53, 1-10.

17. Dornbusch R. 2002. A Primer on Emerging-Market Crises. *Preventing Currency Crises in Emerging Markets.* University of Chicago Press, 16, 743-754.

18. Duca J.V., Muellbauer J., and A. Murphy. 2010. Housing markets and the financial crisis of 2007-2009: Lessons for the future. *Journal of Financial Stability,* 6, 203-217.

19. Durmaz N. 2011. Housing Prices and Fundamentals: The Role of a Supply Shifter. Unpublished manuscript. *Munich Personal RePEc Archive.*

20. Ellis L. 2010. The housing meltdown: Why did it happen in the United States? *International Real Estate Review,* 13, 351-394.

21. Enders W. 2004. *Applied Econometric Time Series.* John Wiley.

22. Engle R.F., and C.W.J. Granger. 1987. Co-integration and error correction: Representation, estimation and testing. *Econometrica,* 55, 251-276.

23. Follain J.R., and R.M. Dunsky. 1997. The Demand for Mortgage Debt and the Income Tax. *Journal of Housing Research*, 8, 155-199.

24. Gimeno R., and C.M. Carrascal. 2006. The Interaction Between House Prices and Loans for House Purchase: The Spanish Case. *Banco de Espana Research Paper*, No: WP-0605.

25. Glennon D. 1989. Estimating the Income, Price, and Interest Elasticities of Housing Demand. *Journal of Urban Economics,* 25, 219-229.

26. Hausman J., and D. McFadden. 1984. Specification tests for the multinomial logit model. *Econometrica,* 52, 1219-1240

27. Hisarciklilar M., Ozturk H., and D.G. Karakas. 2011. Sovereign Risk Ratings: Biased Towards Developed Countries?, *Emerging Markets Finance and Trade,* 47, 69-87.

28. Hofmann B. 2001. The Determinants of Private Sector Credit in Industrialised Countries: Do Property Prices Matter?. *BIS,* No: WP-108.

29. Jaramillo L., and C.M. Tejada. 2011. Sovereign Credit Ratings and Spreads in Emerging Markets: Does Investment Grade Matter? *IMF Working Paper Series*, No: 11/44.

30. Johansen S. 1988. Statistical Analysis of Cointegrated Vectors. *Journal of Urban Economic Dynamics and Control*, 12, 231-254.

31. Kaminsky G., and S.L. Schmukler. 2002. Emerging Markets Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns? *World Bank Economic Review,* 16, 171-195.

32. Kaufman R.K., and J.C. Cutler. 2001. Oil Production in the Lower 48 States: Economic, Geological, and Institutional Determinants. *Energy Journal,* 22, 27-49.

33. Kenny G. 1999. Modelling the Demand and Supply Sides of the Housing Market: Evidence from Ireland. *Economic Modelling,* 16, 389-409.

34. Kristoufek L. 2013. Measuring correlations between non-stationary series with DCCA coefficient. *Physica. A submitted, arXiv*:1310.3984.

35. Larraín G., Reisen H., and J. von Maltzan. 2011. Emerging Market Risk and Sovereign Credit Ratings. *OECD Development Center*, Working Paper No:124.

36. Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis.* Berlin: Springer Verlag.

37. Lütkepohl H., Breitung J., and R. Brüggemann. 2004. New Introduction to Multiple Time Series Analysis. *Applied Time Series Econometrics.* Cambridge University Press.

38. Meese R., and N. Wallace. 1994. Testing the Present Value Relation for Housing Prices: Should I Leave My House in San Francisco?. *Journal of Urban Economics,* 35, 245-266.

39. Pesaran H.M., and Y. Shin. 2002. Long-Run Structural Modelling. *Taylor and Francis Journals,* 21, 49-87.

40. Phillips P.C.B. 1986. Understanding Spurious Regressions in Econometrics. *Journal of Econometrics,* 33, 311-340.

41. Podobnik B. and H. Stanley. 2008. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters,* 100, 084102.

42. Poterba J. 1984. Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach. *Quarterly Journal of Economics*, 99, 729-752.

43. Poterba J. 1992. Taxation and Housing: Old Questions, New Answers. *American Economic Review*, 82, 237-242.

44. Rowland P. 2004. Determinants of Spread, Credit Ratings and Creditworthiness for Emerging Market Sovereign Debt: A Follow-Up Study Using Pooled Data Analysis. *Banco de la Republica (Columbia).*

45. Rowland P., and J.L. Torres. 2004. Determinants of Spread and Creditworthiness for Emerging Market Sovereign Debt: A Panel Data Study. *Banco de la República (Columbia).*

46. Saikkonen P., and H. Lutkepohl. 2000. Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis*, 21, 435-456.

47. Schwab R.M. 1983. Real and Nominal Interest Rates and the Demand for Housing. *Journal of Urban Economics,* 13, 181–195.

48. Skaarup M., and S. Bødker. 2009. House Prices in Denmark: Are They Far from Equilibrium?. *Danish Ministry of Finance,* Working Paper No:21.

49. Solomon B.D. 2010. Are There Any Spillovers between Household and Firm Financing Frictions? A Dynamic General Equilibrium Analysis. PhD Thesis. Université de Montréal: Canada.

50. Somerville C.T. 1999. Residential Construction Costs and the Supply of New housing: Endogeneity and Bias in Construction Cost Indexes. *Journal of Real Estate Finance and Economics*, 18, 43-62.

51. Stroebel J., and M. Floetotto. 2012. Government Intervention in the Housing Market: Who Wins, Who Loses?. *University of Chicago*, Working Paper.

52. Tsay R.S. 2005. *Analysis of Financial Time Series.* John Wiley.

53. Topel R., and S. Rosen. 1988. Housing Investment in the United States. *Journal of Political Econ*omy, 96, 718-40.

54. Valverde S.C., and F.R. Fernandez. 2010. The Relation between Mortgage Markets and House Prices: Does Financial Instability Make the Difference?. *Federal Reserve Bank of Atlanta CenFIS,* No: WP 10-02.

55. Wilcox R.R., and J. Muska. 2001. Inferences about correlations when there is heteroscedasticity. *British Journal of Mathematical and Statistical Psychology,* 54, 39-47.

56. Wooldridge J.M. 2002. Econometric Analysis of Cross Section and Panel Data. *MIT Press.* Cambridge, MA.

57. Zebende G. 2011. DCCA cross-correlation coefficient: Quantifying level of cross-correlation. *Physica A,* 390, 614-618.

58. Zimmerman D.W., Zumbo B.D., and R.H. Williams. 2003. Bias in estimation and hypothesis testing of correlation. *Psicologica,* 24, 133-158.

7 Tables

Table 2.1: Data Descriptions and Sources

Abbreviation	Data	Description	Frequency	Source
cr	Credit Volume	Real value of mortgage credit volume	Monthly	BDDK
per	Housing Permits (Starts)	Number of construction permits	Monthly	TUIK
y	Disposable Income	Real gross domestic product per capita	Quarterly	TUIK
i	Nominal Interest Rate	Average of mortgage rates in the market	Monthly	CBRT
π^e	Expected Inflation	Official expected inflation rate	Monthly	CBRT
ri	Real Interest Rate	Real value of average mortgage rate	Monthly	CBRT
p	House Price	Real house prices driven by house price index	Monthly	Reidin
cc	Construction Cost	Real value of construction cost index	Quarterly	TUIK

Table 2.2: Unit Root Test Results (ADF) (H_0: Non-stationary)

Variable	Deterministic Terms	Lags	Test Value	Critical Levels		
				1%	5%	10%
cr	None	2	2.41	-2.61	-1.95	-1.61
Δcr	None	2	-1.94	-2.61	-1.95	-1.61
per	None	2	0.11	-2.61	-1.95	-1.61
Δper	None	2	-7.26	-2.61	-1.95	-1.61
p	Trend and Intercept	2	-2.13	-4.14	-3.50	-3.18
Δp	Trend and Intercept	2	-3.75	-4.14	-3.50	-3.18
y	Trend and Intercept	2	-2.79	-4.14	-3.50	-3.18
Δy	Trend and Intercept	2	-4.59	-4.14	-3.50	-3.18
cc	Trend and Intercept	2	-2.34	-4.14	-3.50	-3.18
Δcc	Trend and Intercept	2	-4.17	-4.14	-3.50	-3.18
π^e	Trend and Intercept	2	-2.78	-4.14	-3.50	-3.18
$\Delta \pi^e$	Trend and Intercept	2	-3.57	-4.14	-3.50	-3.18
i	Trend and Intercept	2	-1.09	-4.14	-3.50	-3.18
Δi	Trend and Intercept	2	-3.39	-4.14	-3.50	-3.18
ri	Trend and Intercept	2	-0.90	-4.14	-3.50	-3.18
Δri	Trend and Intercept	2	-3.61	-4.14	-3.50	-3.18

Table 2.3: Unit Root Test Results (KPSS) (H_0: Stationary)

Variable	Deterministic Terms	Bandwidth	Test Value	Critical Levels		
				1%	5%	10%
cr	Intercept	2	1.95	0.74	0.46	0.35
Δcr	Intercept	2	0.15	0.74	0.46	0.35
per	Intercept	2	0.78	0.74	0.46	0.35
Δper	Intercept	2	0.03	0.74	0.46	0.35
p	Trend and Intercept	2	0.36	0.22	0.15	0.12
Δp	Trend and Intercept	2	0.08	0.22	0.15	0.12
y	Trend and Intercept	2	0.33	0.22	0.15	0.12
Δy	Trend and Intercept	2	0.11	0.22	0.15	0.12
cc	Trend and Intercept	2	0.29	0.22	0.15	0.12
Δcc	Trend and Intercept	2	0.08	0.22	0.15	0.12
π^e	Trend and Intercept	2	0.52	0.22	0.15	0.12
$\Delta \pi^e$	Trend and Intercept	2	0.13	0.22	0.15	0.12
i	Trend and Intercept	2	0.29	0.22	0.15	0.12
Δi	Trend and Intercept	2	0.12	0.22	0.15	0.12
ri	Trend and Intercept	2	0.33	0.22	0.15	0.12
Δri	Trend and Intercept	2	0.10	0.22	0.15	0.12

Table 2.4: Saikkonen & Lutkepohl Cointegration Test Results

H_0	LR Test Results	p Value	Critical Levels		
			1%	5%	10%
r = 0	241.31	0.0000	163.38	152.59	147.04
r = 1	141.21	0.0009	129.39	119.77	114.84
r = 2	85.55	0.1174	99.40	90.95	86.64
r = 3	81.14	0.0013	73.42	66.13	62.45
r = 4	41.95	0.1066	51.45	45.32	42.25
r = 5	23.26	0.2017	33.50	28.52	26.07
r = 6	11.50	0.2193	19.71	15.76	13.88
r = 7	0.12	0.9915	9.73	6.79	5.47

Table 2.5: Diagnostic Test Results

Type of Test	Test Results	p Value
LM Test for Autocorrelation	160.2502	0.0282
ARCH-LM Test (Δcr_t)	0.8084	0.6675
ARCH-LM Test(Δper_t)	3.1883	0.2031
Jarque-Bera Test (Δcr_t)	5.3641	0.0684
Jarque-Bera Test (Δper_t)	1.0995	0.5771

Table 2.6: Correlation of Impulse Responses in the Housing Market

Correlation of		**to shocks on**					
the responses of	Horizon	p & y	p & π^e	p & i	y & π^e	y & i	π^e & i
cr	5	-0.8609	-0.9664	0.9877	0.9434	-0.9256	-0.9937
	10	-0.8546	-0.9408	0.9748	0.9722	-0.9459	-0.9923
	20	-0.8378	-0.9286	0.9698	0.9148	-0.9035	-0.9907
	50	-0.8488	-0.9279	0.9697	0.8955	-0.8938	-0.9905
per	Horizon	p & cc	p & ri	cc & ri			
	5	-0.0025	-0.8191	-0.5713			
	10	-0.0950	-0.8271	-0.4803			
	20	-0.1295	-0.8329	-0.4399			
	50	-0.1518	-0.8359	-0.4144			

Table 2.7: Structural VECM Forecast Error Variance Decomposition

Proportions of		**accounted for by**							
forecast error in	Horizon	Δcr	Δper	Δp	Δy	Δcc	$\Delta \pi^e$	Δi	Δri
Δcr	1	0.00	0.01	0.68	0.12	0.18	0.00	0.00	0.00
	5	0.00	0.01	0.88	0.04	0.03	0.02	0.02	0.00
	20	0.00	0.00	0.72	0.20	0.00	0.04	0.03	0.00
	50	0.00	0.00	0.71	0.23	0.00	0.03	0.03	0.00
Δper	1	0.00	0.00	0.16	0.06	0.64	0.06	0.07	0.01
	5	0.00	0.00	0.41	0.12	0.27	0.08	0.05	0.06
	20	0.00	0.00	0.47	0.11	0.24	0.07	0.04	0.06
	50	0.00	0.00	0.56	0.09	0.20	0.06	0.04	0.05
Δp	1	0.00	0.00	0.86	0.06	0.02	0.03	0.03	0.02
	5	0.00	0.00	0.92	0.01	0.01	0.02	0.02	0.02
	20	0.00	0.00	0.92	0.03	0.01	0.01	0.01	0.02
	50	0.00	0.00	0.92	0.04	0.01	0.01	0.01	0.01
Δy	1	0.00	0.01	0.04	0.61	0.11	0.15	0.06	0.00
	5	0.00	0.01	0.02	0.64	0.13	0.15	0.05	0.00
	20	0.00	0.01	0.04	0.66	0.08	0.13	0.05	0.04
	50	0.00	0.00	0.04	0.69	0.06	0.12	0.05	0.05
Δcc	1	0.00	0.00	0.00	0.71	0.04	0.06	0.14	0.03
	5	0.00	0.00	0.00	0.71	0.04	0.06	0.14	0.03
	20	0.00	0.00	0.00	0.71	0.04	0.06	0.14	0.03
	50	0.00	0.00	0.00	0.71	0.04	0.06	0.14	0.03
$\Delta \pi^e$	1	0.00	0.00	0.77	0.01	0.05	0.09	0.02	0.06
	5	0.00	0.01	0.57	0.06	0.05	0.17	0.07	0.06
	20	0.00	0.01	0.53	0.10	0.06	0.17	0.07	0.06
	50	0.00	0.01	0.53	0.10	0.06	0.17	0.07	0.06
Δi	1	0.01	0.14	0.05	0.07	0.12	0.33	0.00	0.28
	5	0.00	0.06	0.33	0.10	0.15	0.21	0.03	0.12
	20	0.00	0.03	0.36	0.23	0.11	0.15	0.04	0.07
	50	0.00	0.03	0.42	0.20	0.10	0.14	0.04	0.07
Δri	1	0.02	0.14	0.08	0.08	0.33	0.18	0.01	0.17
	5	0.00	0.04	0.44	0.21	0.13	0.09	0.04	0.04
	20	0.00	0.03	0.32	0.37	0.12	0.09	0.05	0.02
	50	0.00	0.03	0.32	0.37	0.12	0.09	0.05	0.02

Table 3.1: S&P Rating System and Schema of Linear Scales

	Credit Rating Description	Rating	Watch	Linear Transformation		
				Scale 1	Scale 2	Scale 3
Investment Grade	High Grade	AA-	Positive	46	30	16
			Stable	45	30	
			Negative	44	30	
	Upper Medium Grade	A+	Positive	43	29	15
			Stable	42	29	
			Negative	41	29	
		A	Positive	40	28	14
			Stable	39	28	
			Negative	38	28	
		A-	Positive	37	27	13
			Stable	36	27	
			Negative	35	27	
	Lower Medium Grade	BBB+	Positive	34	26	12
			Stable	33	25	
			Negative	32	24	
		BBB	Positive	31	23	11
			Stable	30	22	
			Negative	29	21	
		BBB-	Positive	28	20	10
			Stable	27	19	
			Negative	26	18	
Speculative Grade	Substantial Credit Risk	BB+	Positive	25	17	9
			Stable	24	16	
			Negative	23	15	
		BB	Positive	22	14	8
			Stable	21	13	
			Negative	20	12	
		BB-	Positive	19	11	7
			Stable	18	10	
			Negative	17	9	
	High Credit Risk	B+	Positive	16	8	6
			Stable	15	7	
			Negative	14	6	
		B	Positive	13	5	5
			Stable	12	4	
			Negative	11	3	
		B-	Positive	10	2	4
			Stable	9	2	
			Negative	8	2	
	Very High Credit Risk	CCC+	Positive	7	1	3
			Stable	6	1	
			Negative	5	1	
		CCC	Positive	4	1	2
			Stable	3	1	
			Negative	2	1	
	Default	SD		1	1	1

Table 3.2: Rating Migration

Transition in Notches	Scale 1	Scale 2	Scale 3
-5	0	0	0
-4	0	0	0
-3	1	1	0
-2	3	3	0
-1	11	8	4
0	240	246	275
1	24	27	32
2	13	13	1
3	18	13	0
4	1	1	0
5	1	0	0
Percentage of Transition			
Within 0 Notch*	76.9%	78.8%	88.1%
Within 1 Notch**	88.1%	90.1%	99.7%
Within 2 Notches***	93.3%	95.2%	100.0%

Note: S&P's credit rating transition in notches for different scales.
1-) (+) and (-) signs in notches present the upgrades and downgrades in credit ratings respectively.
2-) *, ** and *** indicate 0, $\pm$ 1 and $\pm$ 2 transitions in credit ratings respectively.

Table 3.3: Data Descriptions and Sources

Data	Description	Frequency	Source
Budget Balance	General Government Budget Performance	Quarterly	Bloomberg
GDP per capita	Per capita gross domestic product in US dollars (Nominal)	Yearly	OECD
Current Account	Current Account Deficit	Quarterly	OECD
Inflation	Change of Consumer Prices over the Period of Previous Year	Quarterly	IMF
External Debt	Countries' total consolidated debt positions in US dollars	Quarterly	World Bank
Export	Goods, Value of Exports in US dollars	Quarterly	IMF
Reserves	Total Reserves excluding Gold in US dollars	Quarterly	IMF
Unemployment	Unemployment Rate of Countries	Quarterly	IMF
Exchange Rate	Change the level of National Currency per U.S. Dollar	Quarterly	IMF
Governance Indicators	Average of Governance Indicators	Yearly	World Bank
G1	Control of Corruption	Yearly	World Bank
G2	Government Effectiveness	Yearly	World Bank
G3	Political Stability and Absence of Violence/Terrorism	Yearly	World Bank
G4	Regulatory Quality	Yearly	World Bank
G5	Rule of Law	Yearly	World Bank
G6	Voice and Accountability	Yearly	World Bank

Table 3.4: Pooled OLS Estimation Results

	Scale 1	Scale 2	Scale 3
Budget Balance/GDP	0.3708***	0.3743***	0.1145***
	(5.09)	(5.26)	(4.62)
GDP per capita	0.0016***	0.0016***	0.0005***
	(5.79)	(6.19)	(5.50)
Current Account/GDP	-0.1857**	-0.1849**	-0.0561**
	(-2.47)	(-2.51)	(-2.19)
Inflation	-0.1004***	-0.1003***	-0.0353***
	(-2.88)	(-2.94)	(-2.97)
Governance Indicators	15.1984***	12.8389***	5.1272***
	(15.78)	(13.65)	(15.66)
External Debt/Export	-3.5664***	-3.3184***	-1.1731***
	(-14.28)	(-13.60)	(-13.81)
Reserves/GDP	16.3627***	13.3317***	5.5347***
	(25.53)	(21.29)	(25.39)
Unemployment	-0.0572	-0.0160	-0.0152
	(-1.45)	(-0.41)	(-1.14)
GDP Growth	-0.0495	-0.0640	-0.0359**
	(-0.94)	(-1.24)	(-2.00)
Exchange Rate	-4.4684	-3.7789	-1.1371
	(-1.40)	(-1.22)	(-1.05)
Constant	26.7824***	18.9288***	9.9062***
	(26.71)	(19.32)	(29.05)
R^2	0.8379	0.8042	0,8318
Wooldridge Test	172.086	281.971	55.620

Note: Estimation results for different scaled dependent variables. The t statistics are in parentheses.
1-) * indicates significance at 10 percent level,
2-) ** indicates significance at 5 percent level,
3-) *** indicates significance at 1 percent level.

Table 3.5: Pooled OLS Estimation Results with AR(1) Disturbance Term

	Scale 1	Scale 2	Scale 3
Budget Balance/GDP	0.1907**	0.2083**	0.0582*
	(1.98)	(2.43)	(1.76)
GDP per capita	0.0020***	0.0019***	0.0006***
	(5.36)	(5.35)	(4.91)
Current Account/GDP	0.0093	-0.0008	0.0096
	(0.13)	(-0.01)	(0.39)
Inflation	-0.1302***	-0.1142***	-0.0493***
	(-3.48)	(-3.65)	(-3.88)
Governance Indicators	10.0266***	6.8210***	3.3114***
	(8.03)	(5.67)	(7.66)
External Debt/Export	-2.4567***	-1.6108***	-0.7631***
	(-8.24)	(-6.39)	(-7.44)
Reserves/GDP	9.8405***	5.8182***	3.1763***
	(12.74)	(8.84)	(12.04)
GDP Growth			-0.0133
			(-0.83)
Constant	25.8250***	17.6111***	9.6668***
	(25.88)	(18.71)	(26.83)
R^2	0.5414	0.3901	0.5429
ρ	0.6969	0.7854	0.7127

Note: Estimation results for different scaled dependent variables. The t statistics are in parentheses.
1-) * indicates significance at 10 percent level,
2-) ** indicates significance at 5 percent level,
3-) *** indicates significance at 1 percent level.

Table 3.6: Panel Specific Models with AR(1) Disturbance Term

	Scale 1		Scale 2		Scale 3	
	(RE)	(FE)	(RE)	(FE)	(RE)	(FE)
Budget Balance/GDP	0.1660**	0.1962***	0.1570**	0.1727***	0.0547*	0.0655**
	(2.06)	(2.76)	(2.33)	(2.69)	(1.94)	(2.53)
GDP per capita	0.0040***	0.0035***	0.0035***	0.0027***	0.0013***	0.0011***
	(10.15)	(9.80)	(10.21)	(7.95)	(9.33)	(8.77)
Current Account/GDP	-0.0322	-0.0214	-0.0412	-0.0302	-0.0053	-0.0028
	(-0.61)	(-0.49)	(-0.99)	(-0.82)	(-0.28)	(-0.18)
Inflation	-0.0233	0.0052	-0.0244	0.0077	-0.0162	-0.0068
	(-0.81)	(0.14)	(-1.04)	(0.26)	(-1.6)	(-0.52)
Governance Indicators	4.7732**	5.6857**	3.3226*	5.3579**	1.6085**	2.2116**
	(2.26)	(2.30)	(1.66)	(2.19)	(2.26)	(2.55)
External Debt/Export	-0.0403	0.7135***	0.1420	0.5844***	-0.0195	0.2344***
	(-0.16)	(3.17)	(0.71)	(3.09)	(-0.22)	(2.77)
Reserves/GDP	2.4800***	0.7414	1.4908***	0.3661	0.9732***	0.3579*
	(3.86)	(1.34)	(2.93)	(0.78)	(4.29)	(1.77)
GDP Growth					-0.0183	-0.0185*
					(-1.52)	(-1.8)
Constant	20.6878***	21.7489***	12.9325***	15.0124***	8.0169***	8.4063***
	(13.99)	(91.86)	(8.79)	(86.01)	(16.06)	(92.03)
R^2(within)	0.6779	0.3190	0.7029	0.2431	0.6720	0.2902
R^2 (between)	0.0470	0.0025	0.0278	0.0002	0.0769	0
R^2 (overall)	0.1686	0.0309	0.1421	0.0365	0.2033	0.0421
ρ	0.7693		0.8216		0.7559	

Note: Estimation results for different scaled dependent variables. The t statistics are in parentheses.
1-) * indicates significance at 10 percent level,
2-) ** indicates significance at 5 percent level,
3-) *** indicates significance at 1 percent level.

Table 3.7: Standard ordered probit with clustered robust standard errors

	Scale 1	Scale 2	Scale 3
Budget Balance/GDP	0.1265***	0.1286***	0.1179**
	(3.01)	(3.02)	(2.51)
GDP per capita	0.0006***	0.0006***	0.0005***
	(3.37)	(3.42)	(3.20)
Current Account/GDP	-0.0877**	-0.0880**	-0.0881**
	(-2.11)	(-2.14)	(-1.96)
Inflation	-0.0110	-0.0143	-0.0126
	(-1.07)	(-1.27)	(-1.24)
Governance Indicators	5.2078***	5.2446***	5.3474***
	(4.91)	(4.86)	(4.53)
External Debt/Export	-1.4375***	-1.4392***	-1.4138***
	(-7.01)	(-6.99)	(-6.10)
Reserves/GDP	6.5998***	6.6430***	6.7069***
	(9.07)	(9.04)	(8.18)
GDP Growth			-0.0302
			(-0.86)
Pseudo R^2	0.3425	0.3496	0.4257
Clusters	8	8	8

Note: Estimation results for different scaled dependent variables. The t statistics are in parentheses.
1-) * indicates significance at 10 percent level,
2-) ** indicates significance at 5 percent level,
3-) *** indicates significance at 1 percent level.

Table 3.8: Random Effects Ordered Probit with clustered robust standard errors

	Scale 1	Scale 2	Scale 3
Budget Balance/GDP	0.1752***	0.1761***	0.1192**
	(3.45)	(3.46)	(2.32)
GDP per capita	0.0022***	0.0022***	0.0023***
	(6.63)	(6.61)	(5.5)
Current Account/GDP	-0.0687***	-0.0689***	0.0467
	(-2.71)	(-2.68)	(0.97)
Inflation	0.0566***	0.0536***	0.0057
	(3.72)	(3.35)	(0.26)
Governance Indicators	6.0582***	6.0585***	4.7780***
	(7.08)	(6.91)	(4.97)
External Debt/Export	-0.8403***	-0.8426***	-0.7365**
	(-4.04)	(-3.99)	(-2.26)
Reserves/GDP	4.8837***	4.9052***	4.3979***
	(6.76)	(6.58)	(4.8)
GDP Growth			-0.0455
			(-1.16)
Clusters	8	8	8

Note: Estimation results for different scaled dependent variables. The t statistics are in parentheses.
1-) * indicates significance at 10 percent level,
2-) ** indicates significance at 5 percent level,
3-) *** indicates significance at 1 percent level.

Table 3.9: Evaluation of Prediction Errors

Prediction Error in Notches	Estimation Model				
	Linear			Ordered Response	
	Pooled OLS	Pooled OLS with AR(1) error terms	RE GLS with AR(1) error terms	Ordered Probit with robust st. errors	RE Ordered Probit with robust st. errors
-10	1	1	1	0	0
-9	0	0	0	9	1
-8	4	1	0	6	1
-7	6	0	0	1	3
-6	3	11	0	7	2
-5	6	6	3	3	6
-4	8	7	8	11	13
-3	13	19	15	29	36
-2	22	14	34	7	13
-1	29	34	55	15	14
0	134	126	57	116	137
1	32	28	64	10	13
2	23	16	30	30	28
3	20	13	17	41	39
4	8	17	18	4	6
5	5	11	7	9	4
6	2	8	2	5	2
7	2	2	3	5	0
8	0	0	0	1	0
9	0	3	1	8	0
10	1	0	3	1	0
Percentage of Predictions					
Correctly Predicted	41.88%	39.38%	17.81%	36.25%	42.81%
Within 1 Notch*	60.94%	58.75%	55.00%	44.06%	51.25%
Within 2 Notches**	75.00%	68.13%	75.00%	55.63%	64.06%
Within 3 Notches***	85.31%	78.13%	85.00%	77.50%	87.50%
Within 4 Notches****	90.31%	85.63%	93.13%	82.19%	93.44%
Root Mean Square Error	2.68	3.08	2.85	3.74	2.64

Note: Prediction error in notches for different estimation models.
1-) (+) sign in notches presents that predicted value is higher than actual value and vice versa for (-) sign.
2-) * indicates percentage of predictions within ± 1 notch, **, *** and **** indicate percentage of predictions within ± 2, ± 3, ± 4 notches respectively.

Table 3.10: Country Specific Prediction Errors

Prediction Error	Brazil	China	India	Indonesia	Mexico	Russia	South Africa	Turkey
-10	0	0	0	0	0	0	0	0
-9	0	0	8	0	0	0	0	1
-8	1	0	4	0	0	0	0	1
-7	0	0	0	1	0	0	0	0
-6	0	3	0	0	1	3	0	0
-5	0	1	0	0	2	0	0	0
-4	1	1	0	5	0	0	0	4
-3	2	7	1	1	6	8	4	0
-2	0	0	3	0	0	0	4	0
-1	1	5	0	1	1	3	2	2
0	15	7	6	15	22	14	24	13
1	0	0	0	0	0	2	5	3
2	9	7	3	7	0	2	0	2
3	4	8	7	5	8	7	1	1
4	3	0	0	0	0	1	0	0
5	1	0	2	2	0	0	0	4
6	1	0	2	1	0	0	0	1
7	0	0	4	1	0	0	0	0
8	0	0	0	1	0	0	0	0
9	1	0	0	0	0	0	0	7
10	0	0	0	0	0	0	0	1
Percentage of Predictions								
Correctly Predicted	37,50%	17,50%	15,00%	37,50%	55,00%	35,00%	60,00%	32,50%
Within 1 Notch*	40,00%	30,00%	15,00%	40,00%	57,50%	47,50%	77,50%	45,00%
Within 2 Notches**	62,50%	47,50%	30,00%	57,50%	57,50%	52,50%	87,50%	50,00%
Within 3 Notches***	77,50%	85,00%	50,00%	72,50%	92,50%	90,00%	100,00%	52,50%
Within 4 Notches****	87,50%	87,50%	50,00%	85,00%	92,50%	92,50%	100,00%	62,50%
Weighted Average of Prediction Errors	0,68	-0,65	-0,90	0,73	-0,28	-0,35	-0,35	1,85

Note: Prediction error in notches for different estimation models.
1-) (+) sign in notches presents that predicted value is higher than actual value and vice versa for (-) sign.
2-) * indicates percentage of predictions within ± 1 notch, **, *** and **** indicate percentage of predictions within ± 2, ± 3, ± 4 notches respectively.
3-) Weighted Average= Prediction Error* Frequency.

8 Appendix

8.1 Appendix A: Derivation Details of the Housing Supply

The Equation (21) in Section 2.2.2 is

$$\frac{\psi}{2}\left(\frac{I_0}{H_0}-\delta\right)-\frac{\xi}{(1-\delta)}\frac{\psi}{2\delta}\left(\left(\frac{I_0}{H_0}\right)^2-\delta^2\right)=\frac{\theta}{(1-\delta)}p_0^H-p_0^I \tag{68}$$

where $\xi=\sum_{t=1}^{\infty}\left(\frac{(1-\delta)}{(1+r)}\right)^t$ and $\theta=\sum_{t=1}^{\infty}\left(\frac{(1-\delta)(1+\pi)}{(1+r)}\right)^t$. Both of the left and right hand sides of Equation (68) is simplified by using basic log-linearization techniques. Main parts of simplifications and calculations are presented here. The detailed calculations are available on request. The left hand side can be rewritten in such a way that more clearly provides to use the log-linearization, i.e.,

$$\frac{\psi}{2}\left(\frac{I_0}{H_0}\right)\left(1-\frac{\xi}{2(1-\delta)}\left(\frac{I_0}{H_0}\right)\right)-\psi\left(1-\frac{\xi\delta}{2(1-\delta)}\right) \tag{69}$$

and then it can be log-linearized around one and the left hand side of Equation (68) returns to

$$2log\left[\frac{\psi}{2}\left(\frac{I_0}{H_0}\right)\left(1-\frac{\xi}{2(1-\delta)}\left(\frac{I_0}{H_0}\right)\right)\right]-2log(2)-log\left[\psi\left(1-\frac{\xi\delta}{2(1-\delta)}\right)\right] \tag{70}$$

Log-linearizing it around one again and expressing in a simpler form makes the left hand side into

$$\left(2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}\right)log\left(\frac{I_0}{H_0}\right)+2log\left(\frac{\left(1-\frac{\xi}{2(1-\delta)}\right)}{4\delta\left(1-\frac{\xi\delta}{2(1-\delta)}\right)}\right) \tag{71}$$

Now, the right hand side of Equation (68) is log-linearized around one and it becomes

$$log\left(\frac{\theta}{(1-\delta)}-1\right)+\frac{1}{\left(1-\frac{(1-\delta)}{\theta}\right)}log(p_0^H)-\frac{1}{\left(\frac{\theta}{(1-\delta)}-1\right)}log(p_0^I) \tag{72}$$

where $\theta_0 > (1-\delta)$. Remembering the functional form of direct investment price, i.e., $p_t^I = ri_t^{\alpha} cc_t^{\gamma}$, the right hand side is finalized

$$log\left(\frac{\theta}{(1-\delta)}-1\right)+\frac{1}{\left(1-\frac{(1-\delta)}{\theta}\right)}log(p_0^H)-\frac{\alpha}{\left(\frac{\theta}{(1-\delta)}-1\right)}log(ri_0)-\frac{\gamma}{\left(\frac{\theta}{(1-\delta)}-1\right)}log(cc_0) \quad (73)$$

Therefore, combining the left and right hand sides in Equation (71) and (73), the representative firm's supply equation with time zero values can be deduced by

$$log\left(\frac{I_0}{H_0}\right) = \Theta + \Gamma_p log(p_0^H) - \Gamma_{ri} log(ri_0) - \Gamma_{cc} log(cc_0) \quad (74)$$

where $\Theta = \frac{1}{2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}} log\left(\frac{\left(\frac{\theta}{(1-\delta)}-1\right)\left(4\delta\left(1-\frac{\xi\delta}{2(1-\delta)}\right)\right)^2}{\left(1-\frac{\xi}{2(1-\delta)}\right)^2}\right)$, $\Gamma_p = \frac{1}{\left(2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}\right)\left(1-\frac{(1-\delta)}{\theta}\right)}$, $\Gamma_{ri} = \frac{\alpha}{\left(2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}\right)\left(\frac{\theta}{(1-\delta)}-1\right)}$, $\Gamma_{cc} = \frac{\gamma}{\left(2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}\right)\left(\frac{\theta}{(1-\delta)}-1\right)}$. All the parameters of Γ's are positive, when $r+\delta < \frac{1}{2}$. In the literature, calibrations in almost all studies employ the risk free rate r and the depreciation rate δ which are $r, \delta \leq 0.10$. So, the supply model in this essay provides wider interval for the parameters. Then, $\xi = \sum_{t=1}^{\infty}\left(\frac{(1-\delta)}{(1+r)}\right)^t = \left[\left(1-\frac{1-\delta}{1+r}\right)^{-1}-1\right] = \frac{1-\delta}{r+\delta} > 2(1-\delta)$, so $\left(2-\frac{1}{\left(\frac{(1-\delta)}{\xi}-\frac{1}{2}\right)}\right) > 0$.

Additionally, $\theta = \sum_{t=1}^{\infty}\left(\frac{(1-\delta)(1+\pi)}{(1+r)}\right)^t = \left[\left(1-\frac{(1-\delta)(1+\pi)}{(1+r)}\right)^{-1}-1\right] = \frac{1-\delta}{\frac{r-\pi}{1+\pi}+\delta} > (1-\delta)$, so $\left(\frac{\theta}{(1-\delta)}-1\right) > 0$ and $\left(1-\frac{(1-\delta)}{\theta}\right) > 0$. Therefore, all the parameters of Γ's in Equation (74) are positive.

8.2 Appendix B: Pearson's Correlation Coefficient and Its Properties

8.2.1 Pearson's Correlation Coefficient

Let X_t and Y_t be two random variables with means μ_x and μ_y and standard deviations σ_x and σ_y, respectively. Pearson's usual (product moment) correlation coefficient, ρ_P, is

defined as

$$\rho_P = \sigma_{xy}/(\sigma_x\sigma_y) = \mathbf{Cov}[X_t, Y_t]/(\sigma_x\sigma_y) \tag{A.1}$$

where $\sigma_{xy} = \mathbf{Cov}[X_t, Y_t] = \mathbf{E}[(X_t - \mu_x)(Y_t - \mu_y)]$. An estimator of ρ_P is given by

$$\widehat{\rho}_P = \frac{\sum_{t=1}^{T} \left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{(T-1)s_x s_y} \tag{A.2}$$

where $s_x^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2$ and $s_y^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)^2$. However, maximum likelihood estimator (MLE) of ρ_P has T in the denominator. Notice that $\widehat{\rho}_P$ is a plug-in and a biased estimator of ρ_P, since $\mathbf{E}[\widehat{\rho}_P] = 1 + \rho_P - \rho_P(1-\rho_P)^2/(2T)$. An approximately unbiased version is

$$\widehat{r} = \widehat{\rho}_P\left[1 + \frac{1 - \widehat{\rho}_P^2}{2T}\right]. \tag{A.3}$$

See Zimmerman, Zumbo and Williams (2003) for further details.

8.2.2 Distribution of Pearson's Correlation Coefficient in General Context

Given a random sample $(X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T)$ from the bivariate normal distribution $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_P)$ with $\sigma_x > 0$, $\sigma_y > 0$ and $|\rho_P| \leq 1$. Let $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_P)$. The log likelihood function of $\mathbf{X} = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_T, Y_T))$ is given by

$$\log p(\mathbf{x}, \theta) = -T\left[\log\left(2\pi\sigma_x\sigma_y\sqrt{1-\rho_P^2}\right)\right] - \tag{A.4}$$

$$\frac{1}{2(1-\rho_P^2)}\left[\frac{\sum_{t=1}^{T}(X_t-\mu_x)^2}{\sigma_x^2} - \frac{2\rho_P\sum_{t=1}^{T}(X_t-\mu_x)(Y_t-\mu_y)}{\sigma_x\sigma_y} + \frac{\sum_{t=1}^{T}(Y_t-\mu_y)^2}{\sigma_y^2}\right]$$

The (unrestricted) MLE of θ is $\widehat{\theta} = \left(\overline{X}, \overline{Y}, \widehat{\sigma}_x^2, \widehat{\sigma}_y^2, \widehat{\rho}_P\right)$ where

$$\begin{aligned}
\widehat{\sigma}_x^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2, \\
\widehat{\sigma}_y^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)^2, \\
\widehat{\rho}_P &= \sum_{t=1}^{T}\frac{\left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{T\widehat{\sigma}_x\widehat{\sigma}_y}.
\end{aligned} \tag{A.5}$$

For simplicity, without loss of generality, we let $U_t = \frac{X_t - \mu_x}{\sigma_x}$ and $V_t = \frac{Y_t - \mu_y}{\sigma_y}$, then $(U_1, V_1), (U_2, V_2), \ldots, (U_T, V_T)$ is a random sample from $N(0, 0, 1, 1, \rho_P)$. Then, we get

$$\widehat{\rho}_P = \frac{\sum_{t=1}^{T}\left(U_t - \overline{U}\right)\left(V_t - \overline{V}\right)}{\sqrt{\sum_{t=1}^{T}\left(U_t - \overline{U}\right)^2\sum_{t=1}^{T}\left(V_t - \overline{V}\right)^2}}. \tag{A.6}$$

Observe that the distribution of $\widehat{\rho}_P$ depends only on ρ_P. When $\rho_P = 0$, we have

$$M_n = \frac{\sqrt{n-2}\widehat{\rho}_P}{\sqrt{1 - \widehat{\rho}_p^2}} \sim t_{T-2}$$

where t_{n-2} is Student's t distribution with $n-2$ degrees of freedom (Bickel and Doksum (1977)). When $\rho_P = 0$ and T is large (i.e., as $T \to \infty$), we have

$$\sqrt{T-2}(\widehat{\rho}_P - \rho_P) \xrightarrow{\mathcal{L}} N(0, (1-\rho_P^2)^2)$$

where $\xrightarrow{\mathcal{L}}$ stands for convergence in law (or distribution). To test for

$$H_o : \rho_P = 0 \text{ versus } H_a : \rho_P \neq 0$$

let $\mathbf{\Theta}_a = \{\theta : \rho_P \neq 0\}$ and $\mathbf{\Theta}_o = \{\theta : \rho_P = 0\}$.Then MLE of θ under $\mathbf{\Theta}_a$ is given in (11) and MLE of θ under $\mathbf{\Theta}_o$ is $\widehat{\theta}_o = \left(\overline{X}, \overline{Y}, \widehat{\sigma}_x^2, \widehat{\sigma}_y^2, 0\right)$.

Logarithm of likelihood ratio test (LRT) statistic becomes

$$\log \lambda(\mathbf{x}) = \log p\left(\mathbf{x}, \widehat{\theta}\right) - \log p\left(\mathbf{x}, \widehat{\theta}_o\right) = \cdots = \frac{-T}{2} \log\left(1 - \widehat{\rho}_P^2\right) \tag{A.7}$$

Thus $\log \lambda(\mathbf{x})$ is increasing in $\widehat{\rho}_P^2$ and LRT rejects H_o for large values of $|\widehat{\rho}_P|$.

8.2.3 Distribution of Pearson's Correlation Coefficient for Bivariate Time Series

Let $\{(X_t, Y_t)\}$ be the bivariate time series,

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \sum_{k=-\infty}^{\infty} C_k \begin{pmatrix} Z_{1,t-k} \\ Z_{2,t-k} \end{pmatrix} \tag{A.9}$$

where $\{(Z_{1,t}, Z_{2,t})\}$ is i.i.d. with mean $(\mu_1, \mu_2)\prime$ and covariance matrix $\sum$, and $C_k = [C_k(i,j)]_{i,j=1}^2$ is a sequence of matrices such that

$$\sum_{k=-\infty}^{\infty} |C_k(i,j)| < \infty,\ i,j = 1,2.$$

Then as $n \longrightarrow \infty$,

$$\hat{\gamma}_X(0) = \hat{\sigma}_X^2 \xrightarrow{p} \gamma_X(0) = \sigma_X^2,$$

$$\hat{\gamma}_Y(0) = \hat{\sigma}_Y^2 \xrightarrow{p} \gamma_Y(0) = \sigma_Y^2 \text{ and}$$

$$\hat{\gamma}_{XY}(0) = \widehat{\mathbf{Cov}}(X,Y) \xrightarrow{p} \gamma_{XY}(0) = \mathbf{Cov}(X,Y),$$

where

$$\hat{\gamma}_X(0) = \hat{\sigma}_X^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2 \text{ with } \bar{X}_n = \frac{1}{n}\sum_{t=1}^{n}X_t \text{ and}$$

$$\hat{\gamma}_Y(0) = \hat{\sigma}_Y^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2 \text{ with } \bar{Y}_n = \frac{1}{n}\sum_{t=1}^{n}Y_t.$$

Furthermore, $\widehat{\rho}_{XY}(0) = \widehat{\rho}_P \xrightarrow{p} \rho_{XY}(0) = \rho_P$, where

$$\widehat{\rho}_{XY}(0) = \frac{\hat{\gamma}_{XY}(0)}{\sqrt{\hat{\gamma}_X(0)\hat{\gamma}_Y(0)}} = \frac{\widehat{\mathbf{Cov}}(X,Y)}{\hat{\sigma}_X\hat{\sigma}_Y}.$$

Therefore, for infinite order moving averages, the estimation of variances, covariance, and correlation for $\{(X_t, Y_t)'\}$ are consistent. Note that $\{Z_t\}$ is white noise which is a special case of this situation, where $\{Z_t\}$ is white noise, if it is stationary with mean 0, and autocorrelation and cross-correlation of $\{Z_{1,t}\}$ and $\{Z_{2,t}\}$ at lag $s > 0$ are all zero. See Brockwell and Davis (2009) for more general results.

If, additionally, the bivariate time series $\{(X_t, Y_t)'\}$ are independent moving averages i.e.,

$$\begin{aligned} X_t &= \sum_{j=-\infty}^{\infty} \alpha_j Z_{1,t-j}, \{Z_{1,t}\} \, is \; i.i.d(0, \sigma_1^2) \text{ and} \\ Y_t &= \sum_{j=-\infty}^{\infty} \beta_j Z_{2,t-j}, \{Z_{2,t}\} \, is \; i.i.d(0, \sigma_2^2), \end{aligned} \tag{A.10}$$

where $\{Z_{1,t}\}$ and $\{Z_{2,t}\}$ are also independent, $\sum_j |\alpha_j| < \infty$, and $\sum_j |\beta_j| < \infty$.

Then,

$$\widehat{\rho}_{XY}(0) = \widehat{\rho}_P \xrightarrow{a} N(0, \frac{1}{n}\sum_{j=-\infty}^{\infty} \rho_X(j)\rho_Y(j))$$

where $\xrightarrow{a}$ stands for "asymptotically distributed as" and $\rho_X(j) = \frac{\gamma_X(j)}{\gamma_{XX}(j)}$ and $\rho_Y(j) = \frac{\gamma_Y(j)}{\gamma_{YY}(j)}$, i.e., $\rho_X(j)$ and $\rho_Y(j)$ are autocorrelations of $\{X_t\}$ and $\{Y_t\}$ at lag j. respectively.

In particular, in the special case of $\{(Z_{1,t}, Z_{2,t})\}$ being white noise, $\widehat{\rho}_{XY}(0) = \widehat{\rho}_P \xrightarrow{a} N(0, 1/n)$.

8.2.4 Correlation Coefficients Written in Matrix Form

An estimator of Pearson's correlation coefficient ρ_P is given by

$$\widehat{\rho}_P = \frac{\sum_{t=1}^{T} \left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{(T-1)s_x s_y} \tag{75}$$

Numerator of $\widehat{\rho}_P$ in (21) can be written as

$$\widehat{\rho}_{P'} = (\mathbf{X}'\mathbf{A})(\mathbf{A}\mathbf{Y}) \tag{76}$$

where $\mathbf{X}' = (X_1, X_2, \ldots, X_T)$ and $\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_T)$, and

$$\mathbf{A} = \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \cdot & \cdot & \cdot & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & \cdot & \cdot & \cdot & -\frac{1}{T} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -\frac{1}{T} \\ -\frac{1}{T} & \cdot & \cdot & \cdot & -\frac{1}{T} & 1-\frac{1}{T} \end{pmatrix}.$$

Here $\mathbf{A}$ is a $T \times T$ matrix, and $\mathbf{X}$ and $\mathbf{Y}$ are $T \times 1$ vectors, with $'$ standing for the transpose

of a vector or matrix. Notice that

$$\mathbf{A}^2 = \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & . & . & . & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & . & . & . & -\frac{1}{T} \\ . & . & . & . & . & . \\ . & . & . & . & . & -\frac{1}{T} \\ -\frac{1}{T} & . & . & . & -\frac{1}{T} & 1-\frac{1}{T} \end{pmatrix}.$$

Hence, $\mathbf{A}$ is also an idempotent matrix. So, the estimate of the Pearson's correlation coefficient in (21) can be written as

$$\widehat{\rho}_P = \frac{\mathbf{X'AY}}{(T-1)\sqrt{\mathbf{X'AX}}\sqrt{\mathbf{Y'AY}}}. \tag{77}$$

A similar formula can be constructed for our new correlation coefficient. Numerator in $\widehat{\rho}_o$ can be written as

$$\widehat{\rho}_{o'} = (T-1)A_{xy} = (\mathbf{CX})'(\mathbf{CY}) = (\mathbf{X'C'})(\mathbf{CY}) \tag{78}$$

where $\mathbf{X'} = (X_1, X_2, \ldots, X_T)$ and $\mathbf{Y'} = (Y_1, Y_2, \ldots, Y_T)$, and

$$\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 & . & . & 0 \\ 0 & -1 & 1 & 0 & . & 0 \\ 0 & 0 & -1 & 1 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & 0 & -1 & 1 \end{pmatrix}.$$

Notice that $\mathbf{C}$ is a $(T-1) \times T$ matrix, and $\mathbf{X}$ and $\mathbf{Y}$ are $T \times 1$ vectors. We compute the product which is a $(T-1)$x$(T-1)$ matrix, as

$$\mathbf{M} = \mathbf{C}' \cdot \mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -1 & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & -1 & 1 \end{pmatrix}.$$

Hence, we can reformulate the new correlation coefficient as

$$\widehat{\rho}_O = \frac{\mathbf{X}'\mathbf{M}\mathbf{Y}}{(T-1)\sqrt{\mathbf{X}'\mathbf{M}\mathbf{X}}\sqrt{\mathbf{Y}'\mathbf{M}\mathbf{Y}}}. \tag{79}$$