

Can Tightness in the Housing Market Help Predict Subsequent Home Price Appreciation? Evidence from the U.S. and the Netherlands¹

Paul E. Carrillo
*George Washington
University*

Erik R. De Wit
*University of Amsterdam
and QuantVille Finance*

William Larson
*George Washington
University*

August 2012

Abstract

This paper assesses the predictive power of variables that measure market tightness, such as seller's bargaining power and sale probabilities, on future home prices. Theoretical insights from a stylized search-and-matching model illustrate that such indicators can be associated with subsequent home price appreciation. The empirical analysis employs data on all residential units offered for sale through a real estate broker in the Netherlands and a large suburb in the Washington, DC area. Individual records are used to construct a quarterly home price index, an index that measures seller's bargaining power, and (quality adjusted) home sale probabilities. Using conventional time-series models we show that current sale probabilities and bargaining power can significantly reduce home price appreciation forecast errors.

Keywords: Forecasting, Home prices, Bargaining power, Time on the market, Information asymmetries

JEL Codes: R30, C53

¹ We would like to thank Bryan Boulier, Ed Coulson, Tara Sinclair, Jeremy Skog, Herman Steckler, Anthony Yezer, and participants at the AREUEA 2012 Midyear Conference and the 2012 ERES Conference for helpful comments and discussions. We are also grateful to MRIS and the Dutch NVM for sharing the data with us. The usual disclaimer applies.

1. Introduction

A large body of literature has documented a strong positive correlation between the rate of change in home prices and the volume of sales in the housing market. Studies suggest alternative (but not necessarily mutually exclusive) explanations for such correlation including down-payment constraints (Stein 1995 and Ortalo-Magne and Rady 1999, 2006), nominal loss aversion (Genesove and Mayer 2001 and Engelhart 2003) and frictions in the search-and-matching process of home buyers and sellers (Berkovec and Goodman 1996, Krainer 2001 and Novy-Marx 2009). Search-and-match markets where the number of buyers is large relative to the number of sellers are said to be "hot" or "tight," are expected to be more liquid, more expensive, and have higher turnover rates and sale volumes than their "cold" counterparts. All of these studies rationalize and carefully document the *contemporaneous* correlation between sales volume, liquidity and home prices.

Home appreciation rates have also shown to be persistent over time (Case and Shiller 1989 and 1990, and Cutler et al. 1991). The observed positive association between past and current home appreciation rates is inconsistent with the predictions of conventional asset pricing models and suggests that home prices may be predictable in the short run.² Given the obvious contemporaneous association between tightness and home prices and the fact that prices seem to be predictable, in this study, we investigate the *predictive* power of variables that measure market tightness on future home prices.

In the U.S. and other developed countries such as the Netherlands, rich data documenting the marketing process of housing sales are generally available. Besides sale prices, these data

² Case and Shiller (1989) and Glaeser and Gyourko (2006) argue that persistence in house price dynamics cannot be explained by fundamentals and may be due to market inefficiencies.

typically contain specific details about each individual transaction such as the list price, marketing time (time on the market) and even details about the bargaining process between buyers and sellers. These data are generally collected by real estate agents in a database system known in the U.S. as Multiple Listing Services (MLS). Although micro-level MLS data are not always available to researchers, some associations of real estate agents compute and publish aggregate statistics such as mean list prices, mean marketing time, the share of transaction below the list price, among many others. It is clear that such statistics provide valuable information about market tightness that may be useful to assess market conditions. It is surprising, however, that such indicators or indices that combine them are not currently being produced systematically in all urban areas to measure the performance of real estate. It is even more surprising that the information in these additional variables is not being used formally to predict the future path of housing prices. In this paper, we show that indicators that combine MLS data could be particularly useful to measure market tightness and to improve predictions about home price appreciation in the short run.

A stylized search-and-matching model is first used to illustrate that indicators that measure market tightness, such as sale probabilities and seller's bargaining power, can be associated with future home price appreciation. The theoretical model developed by Berkovec and Goodman (1996) suggest that changes in housing turnover precede changes in home prices. In their model, buyers and sellers have incomplete information: they observe market transactions but are uninformed about underlying market conditions. Sellers set their reservation values based on their expectations about housing demand; seller's expectations about demand gradually adjust as they gather price information. In the steady state, the flow of buyers and sellers in and out of the market are equal, and buyers and sellers have accurate perceptions about market conditions.

In this setting, a positive shock to buyers (that increases the flow of buyers into the market and thus market tightness) will immediately raise housing demand. Because it takes time for sellers to gather information and adjust their reservation values, a positive shock to buyers will lead to a sharp increase in transactions (faster sales); a gradual increase in seller's reservation values and prices will follow. We add insights from Berkovec and Goodman (1996) into a simple search-and-matching model of the housing market. The matching model is similar in spirit to Novy-Marx (2009) but we allow buyers and sellers to have imperfect information about market conditions. We also allow the ratio of buyers and sellers in the market to affect home seller's bargaining power. That is, bargaining power depends on market tightness to capture the notion that in a tighter (slower) market sellers tend to have higher (lower) levels of market power. The model is used to simulate market outcomes and to illustrate that current sale rates and seller's bargaining power can be associated with future home price appreciation.

The empirical analysis uses residential real estate transaction data from the Netherlands and from Fairfax County, Virginia, a large (380,000 households in 2010) suburban region of the Washington D.C. metropolitan statistical area. The Netherlands data contain all houses and apartments offered for sale through all real estate brokers associated with the Dutch NVM (Dutch Association of Real Estate Brokers and Real Estate Experts) between January 1985 and June 2011. The Dutch NVM data in our sample include more than two million transactions in most regions in the Netherlands and has a market share in the Dutch brokerage market for owner occupied homes of about 70%. Fairfax County real estate transactions were gathered from the local Multiple Listing Services (MLS) and contain more than 300,000 records of listings posted on the system between January 1997 and December 2010. Both datasets include all listings that ended up in a transaction as well as those that expired or were withdrawn from the market and

contain detailed property characteristics, such as the number of bedrooms, bathrooms, age and location, as well as information about list prices, transaction prices and the time that the listing stayed on the market.

The transaction level data are used to construct quarterly aggregate measures of housing conditions in 36 areas in the Netherlands and in 41 zip codes in Fairfax County. Besides home price indices, the indicators include the (quality adjusted) distribution of time on the market, mean difference between list prices and transaction prices, share of transactions below the transaction price, among others. These statistics are used to construct an index that measures seller's bargaining power and to estimate the probability that a home stays on the market for a specific time period. To estimate these indicators we follow the methods proposed by Carrillo (forthcoming) and Carrillo and Pope (2012), respectively.³

The panel of aggregate housing market indicators is used to explore the link between housing sale rates, seller's bargaining power and the rate of change in home prices. In particular, we use conventional time-series models to test if the bargaining power index and the probability of sale have any predictive power to forecast home appreciation rates. Three different approaches are used. First, we estimate a panel autoregressive, distributed lag model (ADL) which assumes that parameters are equal across cross-sectional units and that sale rates and bargaining power are not affected by house prices. Next, we relax the homogeneity assumption and estimate ADL models for each of the cross-sections. Finally, we estimate vector autoregressions (VAR) for each region in our sample. These three approaches are then evaluated on the basis of parameter significance, overall within sample fit, and forecasting performance. We find that lagged sellers' bargaining power and lagged sale probability to be robust empirical determinants of current house price appreciation. Out-of-sample forecasting exercises suggest that these two variables

³ These specific methods are discussed in section (3).

help to forecast house price appreciation one to six quarters ahead versus an AR model baseline. When sellers' bargaining power and sale probabilities are excluded, root mean squared forecasting errors increase by as much as 30 percent.

Given the importance of the real estate market it is not surprising that significant efforts are made on a regular basis to measure current and forecast future housing market outcomes.⁴ The literature has shown that rental values, household income, employment rates, and other variables that measure aggregate macroeconomic conditions can help forecasting home values (Muellbauer and Murphy 1997, Malpezzi 1999, Gallin 2008, Campbell et al. 2009, Rapach and Strauss 2009, Bork and Moller 2012, among others).⁵ Unfortunately, variables that measure macroeconomic conditions are only available for large geographic areas. Hence, forecasts that use methods proposed in the literature can only be made at the aggregate level, typically, at the country or state level. Recent studies have shown significant heterogeneity in home appreciation rates within sub-regions of an urban area (Ferreira and Gyourko 2012, Gleaser et al. 2012, and Guerrieri et al. 2012) which highlights the importance of measuring current and future housing conditions at the local level. The indices and methods developed in this paper show that market tightness indicators can be computed for local markets (even postal codes) and can help forecast appreciation rates in each specific sub-region.

The rest of this paper is organized as follows. Section 2 provides theoretical insights about the relationship between market tightness and home price appreciation. Section 3

⁴ For instance, the level and volatility of housing values, building permits, housing starts, housing inventories, among other indicators, are estimated using home sales records and other administrative data from municipalities and government agencies. Policy makers, investors and lenders are interested in and make regular forecasts about housing market conditions. For example, the U.S. Department of Housing and Urban Development and the U.S. Department of the Treasury produce a monthly Housing Scorecard, including house price forecasts. House price forecasts are also present in HUD's mortgage insurance fund audits and in the Congressional Budget Office's Baseline Economic Forecasts.

⁵ We discuss some of these studies in section 4.

describes the data and constructs housing market aggregate indices. In section 4, we use the aggregate indicators and conventional time series models to forecast home price appreciation. The last section concludes.

2. Conceptual framework

A simple stylized model is developed in this section to illustrate that changes in housing liquidity can precede changes in appreciation rates. The model has the typical structure of search-and-matching models in the literature and is particularly close to the work of Novy-Marx (2009).⁶ A common feature of these models is that selling price and marketing time are simultaneously determined. We modify the classical framework in two ways. First, we add some of the insights provided by Berkovec and Goodman (1996) and Anenberg (2012) and allow buyers and sellers to have imperfect information about market conditions. Second, we allow the ratio of buyers and sellers in the market to affect home seller's bargaining power. These assumptions create dependence between current market tightness and future home price appreciation.

Buyers and sellers are infinitely lived agents who every period, t , search for a potential trading partner. When a buyer and a seller meet, an i.i.d. match value that is specific to the buyer-seller pair is revealed to both of them. The variable ε represents the value of the match between one particular home buyer and one particular home seller. It varies for each buyer-home combination to capture the fact that buyer's preferences for home attributes are heterogeneous. Its distribution is assumed to be known to all agents and is constant across period. That is, from the point of view of buyers and sellers, future values of ε are i.i.d. realizations of a random variable with a known cumulative distribution G . At the time a buyer and a seller meet, they decide

⁶ Search and matching models have been widely used to explain buyers' and sellers' behavior in the housing market (Yinger 1981, Yavas 1992, Horowitz 1992, Albrecht et al. 2007, Novy-Marx 2009, Haurin et al. 2010, Turnbull and Zahirovic-Herber 2011 and Carrillo 2012, among many others).

between trading or continuing searching for a new match next period. If trade occurs, both exit the market forever.

Let m_b and m_s be the measure of active buyers and sellers in the market, respectively. It is assumed that the rate at which sellers (buyers) meet buyers (sellers) depends on market tightness λ , defined as the ratio of buyers to sellers $\lambda = m_b/m_s$. The higher this ratio λ the more (less) likely a seller (buyer) can find a buyer (seller) in a particular period. Let $q_b(\lambda)$ and $q_s(\lambda)$ denote the per-period probability that a buyer and seller find a match, respectively.

The search-and-matching process entails costs to both buyers and sellers. We assume that all agents discount the future using a common discount factor $\beta = 1/(1 + r)$, where r is the discount rate. Moreover, during each meeting, buyers and sellers pay a fixed cost c_b and c_s , respectively.

Denote V_{bt} and V_{st} to the buyer's and seller's value of continuing search, respectively. The surplus from a match is then $\varepsilon - V_{bt} - V_{st}$. Trade will occur only if this surplus is positive. In this case, buyer and sellers split the surplus according to their Nash bargaining powers. Let θ and $1 - \theta$ be the seller's and buyer's bargaining power, respectively.

A seller's value of search V_{st} is the continuation value of walking away from a potential match and it is equal to the discounted expected value of having a chance to find a match in the following period

$$V_{st} = \beta q_s E[\max\{V_{s,t+1} + \theta(\varepsilon - V_{b,t+1} - V_{s,t+1}), V_{s,t+1}\} - c_s] + \beta(1 - q_s)V_{s,t+1}. \quad (1)$$

The maximum operand denotes the optimal seller's decision of matching only when the surplus from the match exceeds her opportunity cost of continuing searching. The buyer's value of search takes a similar form

$$V_{bt} = \beta q_b E[\max\{V_{b,t+1} + (1 - \theta)(\varepsilon - V_{b,t+1} - V_{s,t+1}), V_{b,t+1}\} - c_b] + \beta(1 - q_b)V_{b,t+1}. \quad (2)$$

Given the stationary nature of the economy and the infinite time horizon, we can focus on the steady state and drop the time subscripts (for now). After some manipulation, equations (1) and (2) become

$$V_s = \frac{\beta}{1-\beta} q_s (\theta E[\max\{\varepsilon - V_b - V_s, 0\}] - c_s), \quad (1a)$$

and

$$V_b = \frac{\beta}{1-\beta} q_b ([1 - \theta] E[\max\{\varepsilon - V_b - V_s, 0\}] - c_b). \quad (2a)$$

Notice that it is optimal to buyers and sellers to follow a reservation strategy: accept the match if and only if ε exceeds a reservation threshold ε^r ; that is, a successful match occurs if $\varepsilon > V_b + V_s = \varepsilon^r$. We add (1a) and (1b), replace the optimality condition and obtain

$$r\varepsilon^r = (q_s\theta + q_b(1 - \theta))E[\max\{\varepsilon - \varepsilon^r, 0\}] - (q_b c_b + q_s c_s). \quad (3)$$

The term $E[\max\{\varepsilon - \varepsilon^r, 0\}] = \int_{\varepsilon^r}^{\infty} (\varepsilon - \varepsilon^r) dG(\varepsilon)$ is decreasing in ε^r . Thus, as long as matching costs are small relative to the potential benefits of matching, a unique solution exists. Given that

optimal strategies are time independent, the time it takes a seller to find a match T^s , commonly referred to as time-on-the-market, follows a geometric distribution. That is,

$$Pr\{T^s = x\} = \omega[1 - \omega]^x, \quad (4)$$

where ω is the probability that a seller finds a match in any given period

$$\omega = q_s[1 - G(\varepsilon^r)]. \quad (5)$$

The properties of the continuous-time version of this model have been extensively analyzed by Novy-Marx (2009).

We modify the standard search-and-matching model described above in two ways. First, we allow seller's marketing power to depend on market tightness by assuming that $\theta(\lambda)$ increases with λ . Bargaining power depends on market tightness to capture the notion that in a tighter (slower) market sellers tend to have higher (lower) levels of market power. Second, we assume that it takes time for agents to learn about changes in market conditions. For instance, when the measure of buyers in the economy changes from m_b^0 to m_b^1 , this change is not immediately noticed by buyers and sellers. As in Berkovec and Goodman (1996) we let agents slowly adapt their expectations about this variable as follows

$$m_{bt}^* = \alpha m_{b,t-1} + (1 - \alpha)m_b^1, \quad (6)$$

where $0 \leq \alpha < 1$. At any period t , agents in the market form their expectations about market tightness $\lambda_t^* = m_{bt}^*/m_s$ and use equation (3) to find the match threshold and apply the optimal matching rules described above. Notice that as agents gather information and reach the new steady state, the expectations about market tightness conform reality.⁷

We are interested in analyzing the *path* of adjustment from one steady state to another when the number of buyers in the market changes. In the uninteresting case when $\alpha = 0$, the adjustment to the new steady state is immediate. However, when $\alpha > 0$ buyers and sellers optimal strategies slowly adjust as they gather information about market tightness. Rather than exploring the general properties of the model, we focus on a concrete example. We assume that ε has a normal distribution with mean $u = \$300,000$ and standard deviation $\sigma = \$50,000$. Additional parametric assumptions about $q_s(\lambda)$, $q_b(\lambda)$ and $\theta(\lambda)$ are needed.⁸

The model has been calibrated and used to simulate market outcomes. Optimal matching thresholds (minimum prices) are displayed in the top panel of Figure 1. During the first 20 periods, we let the ratio of buyers to seller's be 1.5 ($\lambda = 1.5$) and agents expectations about market tightness conform reality. In period $t=21$, the economy suffers an external shock that increases λ from 1 to 2. Agents are not immediately aware of this change. In fact, using equation (6) they are able to partially adjust their expectations about λ and use equation (3) to compute the minimum matching threshold $\varepsilon^r(\lambda_t^*)$. In the following periods, as agents collect more information about λ , minimum prices slowly approach the steady state equilibrium. Given our calibrated parameters, this seems to occur after about 20 periods. In other periods when market

⁷ One can modify the model and let buyers and sellers learn about market conditions using Bayesian updating. This is the approach suggested by Anenberg (2012).

⁸ For simplicity, we let $q_s(\lambda) = 1 - \exp(-\rho_s \lambda)$ and $q_b(\lambda) = 1 - \exp(-\rho_b \lambda^{-1})$. The scalar $\rho_s > 0$ measures how market tightness affects the per-period probability that a seller finds a match. All other things equal, higher values of ρ_s increase seller's matching rate. Similarly, the parameter $\rho_b > 0$ denotes how buyers' matching rate is affected by market tightness. Finally, we assume that $\theta(\lambda) = k * \exp(\lambda) / (1 + k * \exp(\lambda))$, where the scalar $k > 0$ measures how seller's bargaining power responds to changes in λ .

tightness is subject to change, slow adjustment in reservations prices follow. For instance, in period $t=51$ after a negative shock decreases λ to 1.5, minimum prices slowly drop to the new steady state. Similarly, once the ratio of buyers of sellers increases to 2.5 in period $t=81$, minimum prices slowly rise.

The probability that a seller finds a match, however, drastically changes after the shock. This occurs because the probability that a seller finds a match, ω , depends directly on the true matching rate $q_s(\lambda)$. That is,

$$\omega_t = q_s(\lambda)[1 - G(\varepsilon^r(\lambda_t^*))]. \quad (7)$$

This is evidenced in the sharp spikes shown in the middle panel of Figure 1 suggesting that changes in demand (shifts in λ) quickly translate to changes in turnover rather than to prices (this point was made by Berkovec and Goodman 1996). The intuition is straightforward. The demand shock is not immediately translated to prices and, as a result, turnover increases. The third panel in Figure 1 computes seller's bargaining power $\theta(\lambda_t^*)$. Notice that bargaining power depends on the agents' perceived market tightness and that, as prices, it slowly adjusts to $\theta(\lambda)$ as agents gather information.

The stylized model developed above provides interesting insights and suggests that changes in housing liquidity can potentially predict changes in home prices. For instance, results shown in the top and middle panel of Figure 1 clearly suggest that when a large and sudden increase (decrease) in $\omega_s(\lambda)$ occurs, one can expect that home appreciation (depreciation) will follow. Furthermore, results in Figure 1 also suggest that seller's bargaining power $\theta(\lambda)$ and

home prices follow a similar path. In the empirical section, econometric methods will be used to test if changes in $\theta(\lambda)$ and $\omega_s(\lambda)$ can help forecast subsequent appreciation rates.

3. Data and variables

The empirical analysis uses residential real estate data from the Netherlands and from Fairfax County, Virginia. Fairfax County is located in northern Virginia and it is part of the Washington, D.C. metropolitan statistical area. This county hosts more than one million residents and more than 380,000 housing units (2010 Census) and it is one of the richest and best-educated counties in the United States.

The Netherlands data contain all houses and apartments offered for sale through all real estate brokers associated to the Dutch NVM (Dutch Association of Real Estate Brokers and Real Estate Experts) between January 1985 and June 2011. The Dutch NVM data in our sample include more than two million transactions in most regions in the Netherlands and has a market share in the Dutch brokerage market for owner occupied homes of about 70%. Fairfax County residential real estate transaction data was gathered from the local Multiple Listing Service (MLS). We collected information from all housing listings that were posted on the MLS between January 1, 1997 and December 31, 2010. Both datasets include all listings that ended up in a transaction as well as those that expired or were withdrawn from the market. The data contain detailed property characteristics, such as the number of bedrooms, bathrooms, age and location, as well as list prices, transaction prices and the time that the listing stayed on the market (time on the market).

Time on the market is measured by the number of days that the MLS listing stays “active” on the market. For units that are sold, we compute marketing time as the difference

between the date when an offer was accepted and the date when the listing was posted. When a listing is withdrawn from the market or it expires without a sale, we compute the time between the initial listing and withdrawal, and treat it as a censored observation. Notice that we analyze the time that a listing stays on the market, which can be different from the total time that the property has been on the market. This occurs because sellers can withdraw the listing for a few days, weeks or even months and then put the property back on the market as a “new listing.” We exclude from our sample listings with unusually high or unusually low listing prices (top and bottom 1 percent during each year), observations that stayed on the market for more than two years, and observations with missing data. After this cleaning process, we are left with about 2.1 million listings in the Netherlands and 284,678 listings in Fairfax County. A variable list and descriptive statistics are available upon request.

The listing data are used to compute several aggregate quarterly indicators that measure market conditions including a) a home price index, b) a seller’s bargaining power index and c) several indices that describe the distribution of marketing time. These aggregate indicators are computed in 36 areas in the Netherlands and in 41 zipcodes in Fairfax County. We describe each one of them below.

Home price index

Using standard hedonic methods we compute a quarterly housing price index. The index is computed for each zip code in Fairfax County and in each region in the Netherlands. Formally, denote p_{itj} to the sale price of home i , in period (quarter) t , in area j , and x_{itj} to the vector of home attributes that describe the housing unit. This vector includes all the variables previously discussed. We assume that log sale prices are defined by the following relationship

$$\log(p_{itj}) = x_{itj}\delta_t + \nu_{tj} + \mu_{itj}, \quad (8)$$

where δ_t is a vector of parameters and μ_{itj} is an unobserved disturbance. Notice that the hedonic coefficients δ_t are allowed to vary in each quarter to capture heterogeneity in housing demand and supply at each point in time but are assumed to be the same across each geographical submarket. The parameters ν_{tj} are time-region fixed effects and measure average differences in log housing prices between each region in a particular period (quarter) and an omitted category (a period-region that is chosen to be the "base").

Equation (8) is estimated using the transaction data described above and Ordinary Least Squares (OLS).⁹ To avoid cluttering the text, we do not report the hedonic coefficients. We should mention, however, that estimates have generally the expected signs, and that these results are available upon request. The estimates of the time-region fixed effects $\hat{\nu}_{tj}$ make our housing price index.

In both the Netherlands and Fairfax County, the housing price index exhibits substantial variation over time and across regions. For example, in Figure 2 we plot the home price appreciation (annual price index change) of a representative zip code of Fairfax County (zip code 22120, the zip code with the largest number of transactions). Home prices in Fairfax evidence a radically different pattern before and after the financial crisis. In the third quarter of 2006 price appreciation peaked at about 30 percent; by 2008, home values experienced a sharp decline. In Figure 3 we display the annual home price appreciation in the Amsterdam region: appreciation rates peaked in 2000 and in 2008.

⁹ Notice that we use the sample of completed transactions. Separate models are estimated in the Netherlands and in Fairfax.

Figure 4 illustrates some of the geographical variation of the price index during the third quarter of 2003 in Fairfax County. Average price appreciation in zip codes located in the north-west areas seem to be substantially lower than in the south-east, suggesting that there is large heterogeneity in home price appreciation rates across regions. This is consistent with the findings of recent papers in the literature (Ferreira and Gyourko 2012, Gleaser et al. 2012, and Guerrieri et al. 2012).

Seller's bargaining power

We measure bargaining power following the methods proposed by Carrillo (forthcoming). Combining list-price, sale-price and time-on-the-market data, he estimates a "heat index" that summarizes housing market conditions and that has a direct economic interpretation. The index measures home seller's bargaining power, θ , and it describes if the housing market is a sellers' (hot) market or a buyers' (cold) market.¹⁰

The index is a consistent estimate of a parameter that measures bargaining power in a structural model of home seller behavior. To gain intuition it is useful to first provide a brief description of the structural model, which is a stylized standard application of search theory to housing. Every period, sellers wait for buyers to visit and inspect their housing units. If a buyer visits, the final sale price is determined and trade may or may not occur. With a fixed known probability $0 < \theta < 1$, the list price is a take-it-or-leave-it offer to the buyer; this probability measures the seller's bargaining power. If trade does not take place, sellers may wait for a potential buyer next period. The list price affects both the rate at which potential buyers arrive

¹⁰ We are not aware of any competing index that measures seller's bargaining power in the academic literature. On the other hand, one can purchase a proprietary "market heat index" of the real estate market (www.marketheatindex.com) that is similar in spirit to the parameter estimated in this paper. Due to its proprietary nature, however, details about this commercial index are not available.

and the final sale price. In particular, a higher list price decreases the likelihood that a buyer arrives but increases the expected sale price. In the steady state sellers optimally pick the list price and reservation value that maximize her expected gains from searching and trade. This stylized model is parameterized to obtain a closed form solution and to facilitate the estimation process.

It is found that the parameter of interest θ can be consistently estimated using aggregate housing sales data as follows

$$\hat{\theta} = \frac{1}{1 + \frac{\hat{d}}{1-\hat{d}} * \left[\exp\left\{\frac{\hat{p}l - \hat{p}m}{\hat{d}} * \phi\right\}\right] * \left\{\frac{1-\hat{d}}{1-\hat{d} + r\hat{T}}\right\}^{1 + \frac{1-\hat{d}}{r\hat{T}}}} . \quad (9)$$

Here \hat{d} is the share of transactions that occurred below the list price, $\hat{p}l$ is the average *log* list-price, $\hat{p}m$ is the average *log* sale-price, and \hat{T} is the average number of days on the market. The parameter r is the daily discount rate (which has been calibrated to $r=0.0001$) and ϕ is a structural coefficient that is not identified by the data and that needs to be calibrated. Within the structural model, ϕ measures how list prices affect buyers' visiting rates. For our practical purposes, higher values of ϕ make our estimate of the seller's bargaining power more responsive to changes in $\hat{p}l - \hat{p}m$. To estimate equation (9), we let $\phi = 8$. This choice allows $\hat{p}l - \hat{p}m$ to account for almost half of the variation of $\hat{\theta}$. We emphasize, however, that the main results of the paper (discussed in the next section) seem notable robust to this normalization.¹¹

¹¹ Clearly, different values of ϕ can affect the level of the heat index. However, index trends over time, and differences in the index between geographic areas seem robust to the choice of ϕ . Researchers who want to give a higher (lower) "weight" to $\hat{p}l - \hat{p}m$ could choose larger (smaller) values of ϕ .

It is important to note that the coefficient of interest θ measures seller's bargaining power in a stationary environment where the time horizon for both buyers and sellers is long. For this reason, $\hat{\theta}$ is computed using aggregate data from the past calendar year (four quarters) rather than using data from just the current quarter. Thus, the heat index estimated at any point in time summarizes housing conditions over the past year. With these considerations, we estimate $\hat{\theta}_{jt}$ for each area j and each quarter t in our sample.

In both the Netherlands and Fairfax County, seller's bargaining power exhibits substantial variation over time and across regions. For example, Figure 4 clearly illustrates a large geographical variation of seller's bargaining power during the third quarter of 2006 within areas in Fairfax County. Figures 2 and 3 plot the index of Fairfax County zip code 20120, and that of the Amsterdam region, respectively. Seller's bargaining power in Fairfax County coincide with popular perceptions about the "heat" of the market: it accelerated in 2000, reached a peaked in 2005 and collapsed after the financial crisis. In Amsterdam, the seller bargaining power index seems to be closely associated with home appreciation rates. In section (4) we test if the bargaining power index can help forecast subsequent home appreciation rates.

Distribution of time on the market

The model developed in the conceptual framework suggests that sudden changes in the probability that a seller finds a match, ω , may precede changes in future appreciation rates. To test these predictions, we need to compute an index that estimates ω in each area-quarter combination in our sample. To achieve this task, we follow the methods proposed by Carrillo and Pope (2012).

Carrillo and Pope use real estate data from Fairfax County to compute (quality adjusted) time on the market distributions and hazard functions for each year during the period 1997 to 2007. In particular, the duration distribution and hazard function during each year in the sample is simulated assuming that housing units have the same characteristics as homes in a base period. The simulation is based on the decomposition methods proposed by DiNardo, Fortin and Lemieux (1996) and the Kaplan-Meier estimator (Kaplan and Meier 1958). Technical details of the method are provided in an appendix.

For each quarter t and each area j , we simulate the distribution of time-on-the-market assuming that the characteristics of housing units remain as those prevalent in the first quarter of 2000 (the base period).¹² We denote this counterfactual distribution as \hat{F}_{jt} . We then estimate $\hat{\omega}_{jt} = \widehat{Pr}\{T_{ij}^s < a\} = \hat{F}_{jt}(a)$, where T_{ij}^s denotes time on the market (in days) and $a = 15$ days.¹³

Figures 2 and 5 clearly illustrate that home sale probabilities exhibit substantial variation over time and across regions in Fairfax County. For instance, in Fairfax County's zip code 20120 the probability that a home sells in less than two weeks in 1998 is almost four times higher than in 2004. As shown in Figure 3, sale probabilities in the Amsterdam region also seem to have increased significantly between 2004 and 2008. In the next section we test if selling rates can help forecast subsequent home appreciation rates.

4. Predicting Home Appreciation Rates

House prices are persistent and forecastable, as Case and Shiller (1989, 1990) demonstrate in their seminal work. This stylized fact is reinforced by Glaeser and Gyourko (2006), who find that a \$1 increase in real house prices predicts a \$0.71 increase the following

¹² Notice that the estimation is separately performed in each area.

¹³ Results are robust if other values of a are chosen (such as $a = 7$, $a = 30$ or $a = 45$ days).

year. Some reasons for this high intertemporal correlation include transaction costs and illiquidity in the housing market (see Muellbauer and Murphy (1997) for some discussion). In general, researchers have shown that it is possible to forecast appreciation with some success using simple, univariate time series techniques because of this persistence in house price.

Building on the simple time series approaches of Case and Shiller, researchers have since found that fundamental economic variables, such as mortgage rates, GDP, incomes, and the unemployment rate influence house prices. Muellbauer and Murphy (1997) find that mortgage market liberalization, demographic shifts, and income explain past U.K. house price growth. Error correction models are popular in this literature, with Malpezzi (1999) showing that house prices correct to location-specific, long-run house price-income ratios, and Gallin (2008) demonstrating the same with house price-rental price ratios. Campbell, Davis, Gallin, and Martin (2009) investigate further the information contained in the price-rent ratio using variance decompositions at the MSA level. In order to develop their forecasts, Campbell et. al estimate region-by-region VAR models with regional and national variables modeled exogenously with respect to the regional housing market.¹⁴

Rapach and Strauss (2009) forecast house prices using a variety of state and regional variables as well, in order to determine which variables help forecast the best in different regions. These variables include state-level house price/income ratios, incomes, unemployment rates, incomes, and population, as well as Census division-level housing starts, permits, and vacancy rates. Building this more data-driven approach, Bork and Moller (2012) estimate a factor model with 122 different macroeconomic variables and find that the factors forecast well relative to rival models.

¹⁴ De Wit et al (2010) include list prices, and sales volume in a similar VAR framework.

While researchers have modeled and forecasted house prices using national, regional, or city-level variables, they have not attempted to do so using (readily available) local housing market variables. The theory presented in this paper suggests that local variables, such as sale probabilities and bargaining power, could be used to help explain and forecast house price appreciation.

The goal of this section is to empirically test if *current* period sellers' bargaining power and home selling rates have a positive effect on future house prices. In order to test these hypotheses, three different approaches are used, generally following the forecasting and evaluation methods of Stock and Watson (1999 and 2003), Campbell et. al (2009), and Rapach and Strauss (2009). First, we demonstrate the basic correlations among the variables using a panel autoregressive, distributed lag model (ADL) across 41 zip codes in Fairfax and 36 regions in the Netherlands. This assumes that parameters are equal across cross-sectional units and that past values of sale rates and bargaining power are pre-determined. Next, we relax the homogeneity assumption and estimate ADL models for each of the cross-sections. Finally, we estimate vector autoregressions (VAR) for each zipcode in Fairfax and each region in the Netherlands. These three approaches are then evaluated on the basis of parameter significance, overall within sample fit, and forecasting performance.

Panel ADL Model:

The theory section describes a model where house prices adjust over time due to information stickiness and time lags in house price adjustment. A partial adjustment model can capture these intertemporal dynamics. Equation (10) presents a general ADL specification, where Δ_{ti} is the year-on-year home price index difference in region i and period t , $\Delta_{ti} = \hat{v}_{ti} - \hat{v}_{t-4,i}$, α_i consists

of deterministic components that are common to each area that are constant over time (zipcode-level fixed effects in Fairfax County and region fixed effects in the Netherlands), and η_t are time fixed effects that capture any trend common to all zipcodes / regions in the sample. In the most simple specification, θ and ω are assumed to be pre-determined and parameters are assumed to be equal across regions.¹⁵

$$\Delta_{it} = \alpha_i + \sum_j \beta_j \Delta_{i,t-j} + \sum_j \gamma_j \theta_{i,t-j} + \sum_j \delta_j \omega_{i,t-j} + \eta_t + \varepsilon_{it} \quad (10)$$

Placing some *a priori* zero-restrictions on gammas and deltas, along with a lag-length of three (chosen based on the Levin, Lin, and Chu (2002) procedure), gives Equation (11),¹⁶

$$\Delta_{it} = \alpha_i + \beta_1 \Delta_{i,t-1} + \beta_2 \Delta_{i,t-2} + \beta_3 \Delta_{i,t-3} + \gamma \theta_{i,t-1} + \delta \omega_{i,t-1} + \eta_t + \varepsilon_{it} \quad (11)$$

where ε_{it} is an i.i.d. disturbance.

Equation (11) is estimated for the Fairfax and Netherlands samples, and results are shown in Table 1a and Table 1b, respectively. For robustness five different specifications have been estimated. All models include zipcode / region fixed effects and home appreciation rate lags. Models (2) and (3) add lagged sellers' bargaining power (θ) and the match rate (ω), respectively.

¹⁵ Note that it is assumed that each region is independent of every other.

¹⁶ This specification maintains the partial adjustment characteristics of the general ADL model, in that a change to theta or omega will have effects multiple periods in the future. The long-run appreciation rate is

$$\Delta_4^* = \frac{\alpha + \gamma\theta + \delta\omega}{1 - \beta_1 - \beta_2 - \beta_3}$$

In both Fairfax and the Netherlands, coefficients are positive and statistically significant. Comparing model (1) to models (2) and (3), we see that model fit increases as θ and ω are individually added.¹⁷ When both variables are added in model (4), both are positive and significant as theory would predict.

As a robustness check to model (4), time period (quarter) fixed effects are included in the model (5).¹⁸ Time period fixed effects take into account any national or regional trends that would affect house prices, such as the MSA-level labor costs, national GDP or interest rates, and construction costs. They capture all variation in home prices that is common in all areas during a specific quarter. The time effects increase the overall fit of the model substantially and reduce to some extent the magnitude of our variables of interest. However, sign and significance are preserved, indicating that the results in model (4) are notably robust.

This panel specification demonstrates that, in general terms, current bargaining power and selling rates are strongly associated with future home appreciation rates, after controlling for lagged appreciation rate, zipcode/region and time fixed effects. This association is strong in both Fairfax and Netherlands samples.

Region-level ADL Model:

The next set of models relaxes the assumption that parameters are equivalent across different zipcodes (in Fairfax) and across different regions (in the Netherlands). It is plausible that the coefficients are different across regions, and a more disaggregated approach would result in a

¹⁷ F-statistics reject the null of equivalent fit in each case.

¹⁸ Time period fixed effects are modeled using a dummy variable for each quarter.

better fit and less information gain from the addition of new variables. Equation (12), given below, is identical to equation (11) except that parameters are allowed to vary over i :

$$\Delta_{it} = \alpha_i + \beta_{1i}\Delta_{i,t-1} + \beta_{2i}\Delta_{i,t-2} + \beta_{3i}\Delta_{i,t-3} + \gamma_i\theta_{i,t-1} + \delta_i\omega_{i,t-1} + \varepsilon_{it} \quad (12)$$

Estimates of the signs and significance levels of the θ s and ω s are tallied and presented in Table 2.¹⁹ Panel A tabulates results for Fairfax County and panel B for the Netherlands. The table shows tabulations of estimates across 41 separate models in Fairfax (by zipcode) and 36 in the Netherlands (by region). Each model is estimated using OLS. Model 1 estimates the same equation shown in the second column of Table 1 in each of the individual areas (zipcodes in Fairfax and regions in the Netherlands) and tabulates the sign and statistical significance of the bargaining power parameter θ . Model 2 estimates the same equation shown in the third column of Table 1 in each zipcode/region and tabulates the sign and statistical significance of the sale probability parameter ω . Model 3 estimates equation shown in the fourth column of Table 1 and tabulates the sign and statistical significance of both θ and ω . In both Fairfax County and the Netherlands, point estimates are generally positive and statistically significant, suggesting that they are robust across regions. F-tests indicate that both omega and theta are jointly significant in the vast majority of cases.

VAR Model:

Finally, we model the relationship between house prices, θ , and ω in a vector autoregression (VAR) framework. A VAR places no *a priori* exogeneity restrictions on any of

¹⁹ The estimates of this model are not presented here and are available upon request.

the dependent variables, instead assuming that each variable can affect each other over time. This is a plausible stochastic specification because, as the theory section suggests, price information affects bargaining power and the probability of a match over time. A VAR model can also be transformed into a vector error correction model given the appropriate restrictions, giving the partial adjustment result from the theory as a special case. The VAR model is presented in Equation (13)

$$Y_{it} = \alpha_i + \beta_{1i}Y_{i,t-1} + \beta_{2i}Y_{i,t-2} + \beta_{3i}Y_{i,t-3} + \varepsilon_{it}, \quad (13)$$

where $Y_{it} = \begin{bmatrix} \Delta_{it} \\ \theta_{it} \\ \omega_{it} \end{bmatrix}$ and $\varepsilon_{it} \sim N(0, \sigma_i^2)$.

Table 3 shows the root mean squared errors (RMSE), both in-sample and over the forecasting period both for Fairfax (Panel A) and the Netherlands (Panel B).²⁰ One quarter forecasts are computed recursively, following Rapach and Strauss (2009).²¹ Results suggest that both θ and ω help to model house prices in sample and forecast better from 2006 to 2010. Each of these variables contains unique and relevant information, as the RMSE falls when both are included in the model.²² F-statistics indicate that the in-sample error is less than the forecasting error in the U.S. sample, suggesting that all models fail parameter constancy. In the U.S., this is likely due to the 2008 financial crisis, as this is part of the forecasting period. Interestingly, in the Netherlands sample, the in-sample error is actually larger than the forecasting error. This could be due to structural breaks in the estimation period, or could simply be due to a fall in

²⁰ In-sample RMSE is computed using the model estimated from the initial sample period until 2005.

²¹ For example, for the 2006q1 forecast in Fairfax County, the model is estimated using data from 1997q1-2005q4; for the 2006q2 forecast, the model is estimated using data from 1997q1-2006q1; and so on.

²² Forecast encompassing tests can also be used to evaluate the relevant information content in rival forecasts. See Chong and Hendry (1986).

structural error variance between the two time periods. F-statistics indicate that each of the three sets of forecasts including θ and/or ω is better than the AR forecast including only the price index appreciation series using the U.S. data. However, with the Netherlands data, only ω improves upon the AR forecast.

Table 4 shows the RMSEs of forecasts at different time horizons. Panel A shows results for Fairfax County. Consistent with the theory presented in this paper, θ and ω help to forecast over short intervals. The three-variable model forecasts are significantly better than the AR forecasts at every time horizon up to eight quarters, though the improvement diminishes as the horizon increases. Results for the Netherlands suggest that all forecasts become worse as the horizon increases. Forecasts including θ are of similar fit to the AR forecast, but the inclusion of ω results in significant improvement. Unlike the U.S. case, ω helps to forecast at all time horizons, and this effect does not diminish with time.²³

5. Conclusions

In this study, we test the predictive power of variables that measure market tightness on future home prices. Theoretical insights from a stylized search-and-matching model suggest indicators that measure market tightness, such as sale probabilities and seller's bargaining power, are associated with future home price appreciation. The empirical analysis uses Multiple Listing Services data from the Netherlands and from Fairfax County, VA, that contain all residential units offered for sale through a real estate broker over a 15 and 20 year period, respectively. The individual records are used to construct quarterly aggregate measures of housing conditions in

²³ As a robustness check, for the U.S. sample we also estimated VAR models that include regional variables such as unemployment rate and interest rates that do not vary within each subregion (for example the same interest rate applies for all postal codes in Fairfax County). We also tested national and Virginia house price changes. When these additional controls are included in the VAR models, RMSE in all models falls, but the forecasting ranking of the models is maintained. These results are not included in the paper and are available upon request.

about 36 regions in the Netherlands and in 41 zip codes in Fairfax County. Besides home price indices, the indicators include an index that measures seller's bargaining power and the (quality adjusted) probability that a home sells in less than 2 weeks. Conventional time-series models have been used to show that observed changes in sale rates and bargaining power can significantly reduce home price appreciation forecast errors.

This research has important implications for both the real estate industry and policy makers. Given the importance of the housing market and the availability of transaction level data, the construction of such indicators on a regular basis for all areas in the U.S., the Netherlands and other developed countries should be a relatively straightforward task that could inform economic agents about local market conditions. Improved forecasts and understanding of current market conditions should be of interest to home buyers and sellers (and their agents) who generally like to be informed about market conditions when setting their optimal marketing strategies and, of course lenders, the PMI industry, and even participants in the derivatives market. Information about seller's bargaining power and housing liquidity could also be relevant to investors and regulators because, it provides information about market risk and, more importantly, it could be a valuable input to predict future home prices.

References

- Albrecht, J., A. Anderson, E. Smith, and S. Vroman (2007), "Opportunistic Matching in the Housing Market," *International Economic Review*, 48, 641-664.
- Anenberg, E. (2012), "Information Frictions and Housing Market Dynamics," unpublished manuscript.
- Berkovec, J.A. and J.L. Goodman (1996), "Turnover as a measure of demand for existing homes," *Real Estate Economics* 24, 421-440.
- Bork, L. and S. V. Møller (2012), "Housing price forecastability: A factor analysis," unpublished manuscript.
- Carrillo, P. (forthcoming), "To Sell or Not to Sell: Measuring the Heat of the Housing Market," *Real Estate Economics*.
- Carrillo, P. (2012), "An Empirical Stationary Equilibrium Search Model of the Housing Market," *International Economic Review*, 53, 203-234.
- Carrillo, P. and J. Pope (2012), "Are Homes Hot or Cold Potatoes? The Distribution of Marketing Time in the Housing Market," *Regional Science and Urban Economics* 42, 189-197.
- Campbell, S. D., M.A. Davis, J. Gallin, and R. F. Martin (2009), "What moves housing markets: A variance decomposition of the rent-price ratio," *Journal of Urban Economics*, 66, 90-102.
- Case, K. and R. Shiller (1989), "The efficiency of the market for single-family homes," *The American Economic Review*, 79, 125-137.
- Case, K. and R. Shiller (1990), "Forecasting prices and excess returns in the housing market," *Real Estate Economics*, 18, 253-273.
- Cutler D. M., J. M. Poterba and L. H. Summers (1991), "Speculative Dynamics," *Review of Economic Studies*, 58, 529-546.
- DiNardo, J., N. Fortin and T. Lemieux (1996), "Labor market institutions and the distribution of wages, 1973–1992: a semiparametric analysis," *Econometrica* 64, 1001–1044.
- De Wit, E., P. Englund, and M. Francke (2010), "Price and Transaction Volume in the Dutch Housing Market," *Tinbergen Institute Discussion Papers* 10-039/2.
- Engelhardt, G.V. (2003), "Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States," *Journal of Urban Economics* 53, 171-195.

- Ferreira, F. and J. Gyourko (2012), "Heterogeneity in Neighborhood-Level Price Growth in the United States, 1993-2009," *American Economic Review*, 102, 134–40.
- Gallin, J. (2008), "The long-run relationship between house prices and rents," *Real Estate Economics*, 4, 635-658.
- Glaeser, E., J. D. Gottlieb, and K. Tobio (2012), "Housing Booms and City Centers," *American Economic Review*, 102, 127–33.
- Genesove, D. and C. Mayer (2001), "Loss aversion and selling behavior: evidence from the housing market," *Quarterly Journal of Economics* 116, 1233-1260.
- Guerrieri, V., D. Hartley and E. Hurst (2012), "Within-City Variation in Urban Decline: The Case of Detroit," *American Economic Review*, 102, 120–26.
- Haurin, D., J. Haurin, T. Naudald and A. Sanders (2010), "List Prices, Sale Prices and Marketing Time: An Application to U.S. Housing Markets," *Real Estate Economics*, 38, 659-685.
- Horowitz, J. (1992), "The Role of the List Price in Housing Markets: Theory and an Econometric Model," *Journal of Applied Econometrics*, 2, 115-129.
- Kaplan, E. L. and P. Meier (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association* 53, 457–481.
- Krainer, J. (2001), "A theory of liquidity in residential real estate markets," *Journal of Urban Economics* 49, 32-53.
- Levin, A, C. Lin, and C. Chu (1992), "Unit root tests in panel data: asymptotic and finite-sample properties," *Journal of Econometrics* 108, 1-24.
- Malpezzi, S. (1999), "A simple error correction model of house prices," *Journal of Housing Economics* 8, 27-62.
- Muellbauer, J. and A. Murphy (1997), "Booms and busts in the U.K. housing market," *The Economic Journal*, 107, 1701-1727.
- Novy-Marx, R. (2009), "Hot and cold markets," *Real Estate Economics* 37, 1-22.
- Ortalo-Magné, F. and S. Rady (1999), "Boom in, bust out: Young households and the housing price cycle", *European Economic Review* 43, 755-766.
- Ortalo-Magné, F. and S. Rady (2006), "Housing market dynamics: on the contribution of income shocks and credit constraints," *Review of Economic Studies* 73, 459-485.
- Rapach, D. and J. Strauss, (2009), "Differences in house price forecastability across U.S. States," *International Journal of Forecasting*, 25, 33-42.

Stein, J. (1995), "Prices and trading volume in the housing market. A model with downpayment constraints," *Quarterly Journal of Economics* 110, 379-406.

Stock, J. and M. Watson (1999), "Forecasting inflation," *Journal of Monetary Economics* 44, 293-335.

Stock, J. and M. Watson (2003), "Forecasting output growth and inflation: the role of asset prices," *Journal of Economic Literature* 41, 788-829.

Turnbull, G. and V. Zahirovic-Herber (2011), "Why Do Vacant Houses Sell for Less: Holding Costs, Bargaining Power or Stigma?" *Real Estate Economics*, 39, 19-43.

Yavas, A. (1992), "A Simple Search and bargaining Model of Real Estate Markets," *Journal of the American Real Estate and Urban Economics Association*, 20, 533-548.

Yinger, J. (1981), "A Search Model of Real Estate Broker Behavior," *American Economic Review*, 71, 591-605.

Appendix

We estimate quality-adjusted time-on-the-market distributions for each quarter-area combination in our sample. The method used to compute the distributions follow Carrillo and Pope (2012) who combine the Dinardo, Fortin and Lemieux (1996) (DFL) with the Kaplan-Meier estimator. This allows the DFL decomposition to work in cases where the dependent variable is subject to random censoring. To keep our exposition self-contained we carefully review the decomposition method.

Let Y be our variable of interest (time a listing stays on the market) and t_0 and t_1 refer to the two mutually exclusive periods (quarters) in each of the areas we analyze. The cumulative probability function of Y in period t_0 is defined as

$$(1A) \quad F(y | T = t_0) = P\{Y \leq y | T = t_0\} = \int F(y | x, T = t_0) h(x | T = t_0) dx,$$

where T is a random variable describing the period from which an observation is drawn and x is a particular draw of observed attributes of individual characteristics from a random vector of housing characteristics X . $F(y | x, T = t_0)$ is the (conditional) cumulative distribution of Y given that a particular set of attributes x have been picked, and $h(x | T = t_0)$ is the probability density of individual attributes evaluated at x . The cumulative probability function of Y in period t_1 is defined similarly.

Suppose we would like to assess how the distribution of Y (marketing time) in period t_1 would look if the individual attributes x (number of bathrooms, bedrooms and age, for example) were the same as in period t_0 (the base quarter). We denote this counterfactual as $F_{t_1 \rightarrow t_0}$ and express it symbolically as²⁴

$$(2A) \quad F_{t_1 \rightarrow t_0} = \int F(y | x, T = t_1) h(x | T = t_0) dx.$$

Using Bayes' rule, DFL recognized that

$$(3A) \quad \frac{h(x | T = t_0)}{h(x | T = t_1)} = \frac{\frac{P(T = t_0 | X = x)}{P(T = t_0)}}{\frac{P(T = t_1 | X = x)}{P(T = t_1)}} = \frac{\frac{P(T = t_0 | X = x)}{1 - P(T = t_0 | X = x)}}{\frac{P(T = t_0)}{1 - P(T = t_0)}} = \tau_{t_1 \rightarrow t_0}(x).$$

One may use expression (3A) to substitute $h(x | T = t_0)$ in equation (2A) and thereby obtain expression (4A).

²⁴ The subscript " $t_0 \rightarrow t_1$ " indicates that the attributes data from period t_0 will be "replaced" by data from period t_1

$$(4A) \quad F_{t_1 \rightarrow t_0}(y) = \int F(y | x, T = t_1) h(x | T = t_1) \tau_{t_1 \rightarrow t_0}(x) dx$$

Notice that this expression differs from equation (1A) only by $\tau_{t_1 \rightarrow t_0}(x)$. DFL refer to $\tau_{t_1 \rightarrow t_0}(x)$ as “weights” that should be applied when computing the counterfactual distribution of our variable of interest. However, given that the weights are unknown, they need to be estimated.

Carrillo and Pope (2012) note that the DFL method described above cannot be directly used in this application because marketing time is subject to random censoring; that is, some properties are not sold and withdrawn from the market. Because the random variable Y (marketing time) is subject to random censoring, the counterfactual distribution can be computed using the Kaplan-Meier estimator, with sampling weights given by $\tau_{t_1 \rightarrow t_0}(x)$.

To be specific, we summarize the estimation algorithm for the counterfactual given that a random sample of N_0 and N_1 observations for periods t_0 and t_1 is available. Notice that in all steps described below the sample includes all censored and non-censored observations.

Step 1: Estimate $P(T = t_0)$ using the share of observations where $T_i = t_0$; that is, compute:

$$\hat{P}(T_i = t_0) = N_0 / (N_0 + N_1).$$

Step 2: Estimate $P(T = t_0 | X = x)$, by estimating a logit model using the pooled data. The dependent variable equals one if $T_i = t_0$, and explanatory variables include the vector of individual attributes x_i .

Step 3: For the subsample of observations where $T_i = t_1$, estimate the predicted values from the

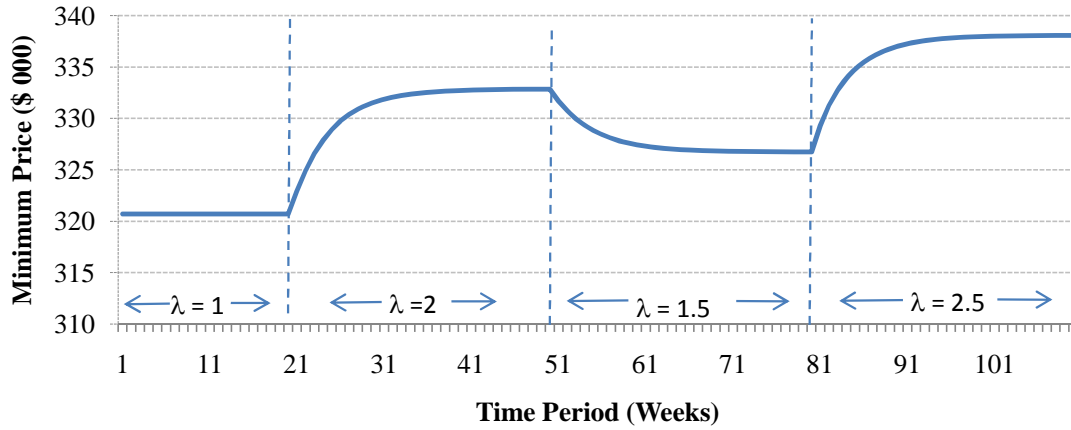
$$\text{logit } \hat{P}(T_i = t_0 | X = x_i) = \frac{\exp\{x_i \hat{\beta}\}}{1 + \exp\{x_i \hat{\beta}\}}, \text{ where } \hat{\beta} \text{ is the parameter vector from the logit}$$

regression. Then, compute the estimated weights $\hat{\tau}_{t_1 \rightarrow t_0}(x)$.

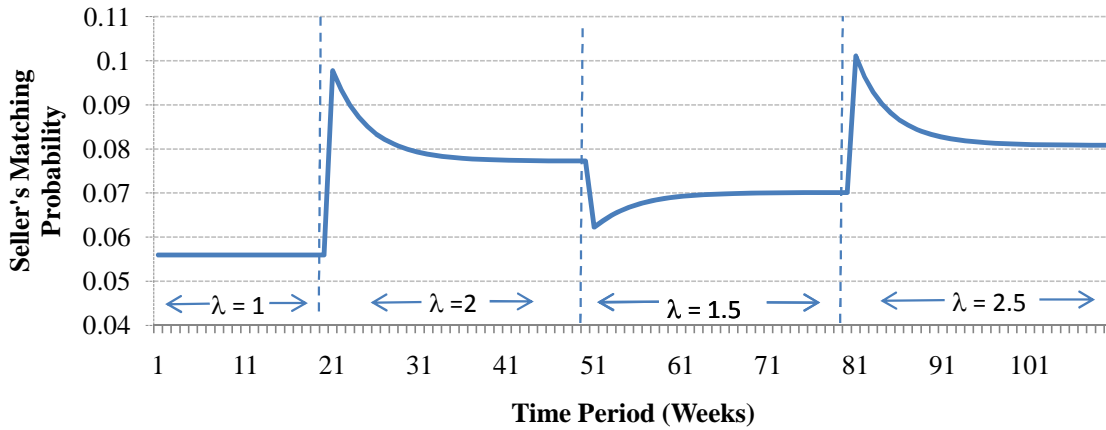
Step 4: For the subsample of observations where $T_i = t_1$, compute a weighted empirical cumulative distribution function using the Kaplan-Meier estimator. Weights are given by $\hat{\tau}_{t_1 \rightarrow t_0}(x_i)$.

Figure 1
Market Response to Changes in Buyer / Seller Ratio (λ)

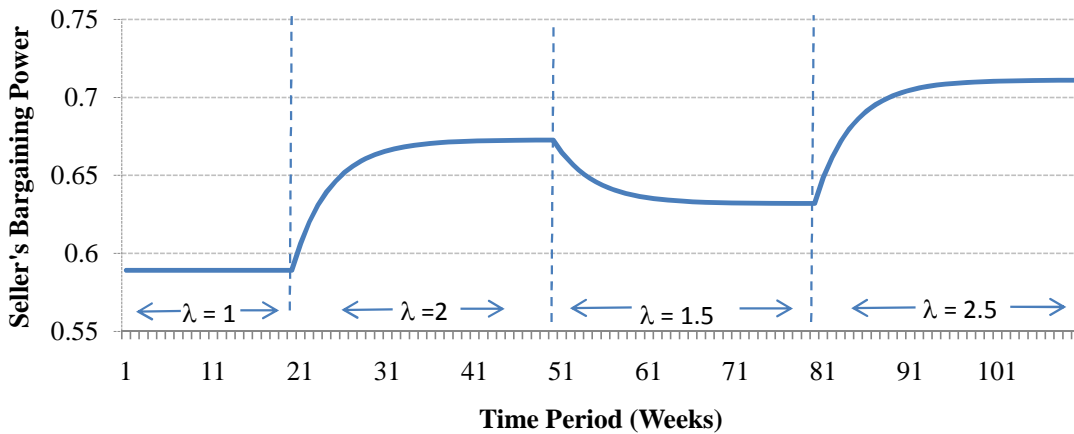
A. Prices



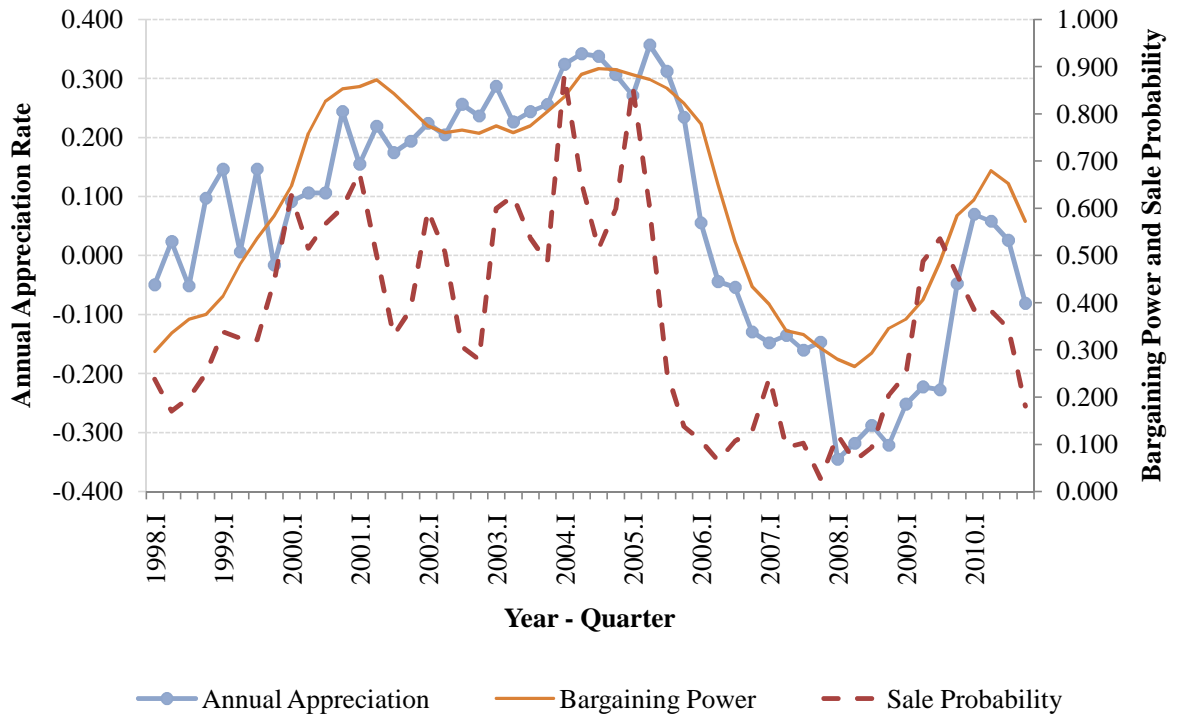
B. Per-Period Seller's Matching Probability



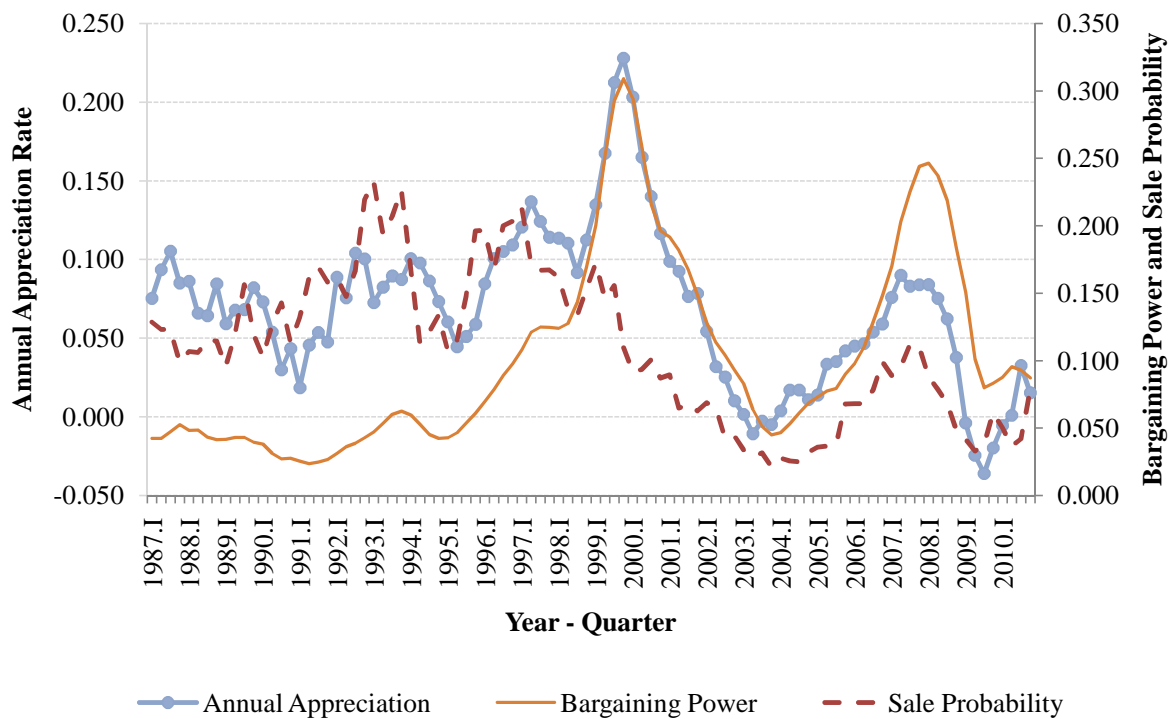
C. Seller's Bargaining Power



**Figure 2: Housing Market Conditions in Fairfax County
- Zip Code 20120 -**

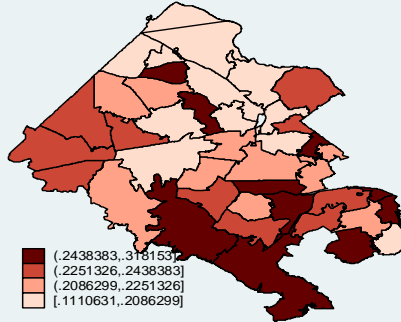


**Figure 3: Housing Market Conditions in the Netherlands
- Amsterdam Region -**

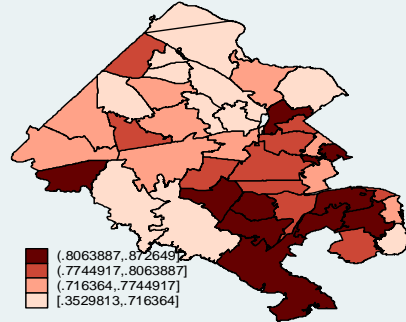


**Figure 4: Appreciation Rates and Bargaining Power
- Fairfax County, Year 2003 -**

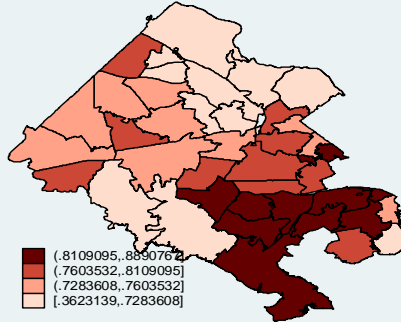
**Annual Appreciation Rates
Year 2003-Q3**



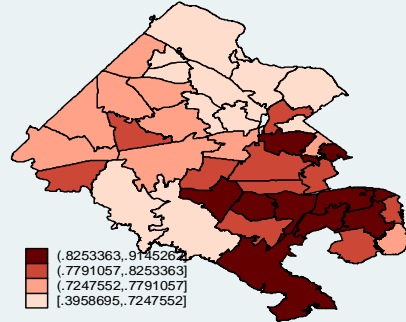
**Sellers' Bargaining Power
Year 2003-Q3**



**Sellers' Bargaining Power
Year 2003-Q2**



**Sellers' Bargaining Power
Year 2003-Q1**



**Figure 5: Appreciation Rates and Sale Probability
- Fairfax County, Year 2003 -**

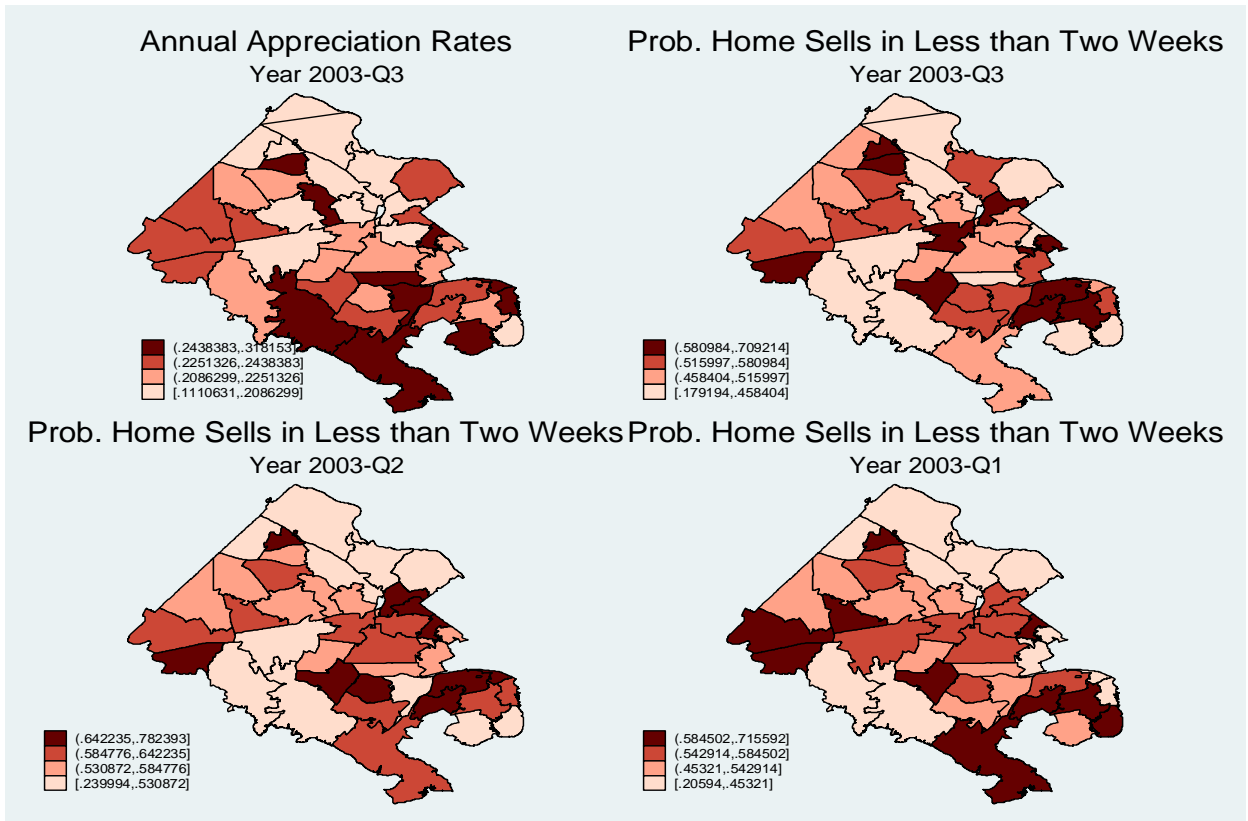


Table 1a**Fairfax County: Panel ADL Estimates****Dependent Variable: $D(t) = [\log \text{ of house prices } p(t) - \log \text{ of house prices } p(t-4)]$**

Variable	Equation				
	[1]	[2]	[3]	[4]	[5]
Bargaining power: θ (t-1)		0.192*** (0.0175)		0.0660*** (0.0178)	0.0725*** (0.0171)
Sale probability: ω (t-1)			0.214*** (0.0104)	0.197*** (0.0113)	0.0193* (0.0103)
D(t-1) = p(t-1) - p(t-5)	0.772*** (0.0226)	0.682*** (0.0235)	0.562*** (0.0229)	0.547*** (0.0232)	0.227*** (0.0232)
D(t-2) = p(t-2) - p(t-6)	0.284*** (0.0278)	0.212*** (0.0277)	0.276*** (0.0252)	0.252*** (0.0259)	0.177*** (0.0240)
D(t-3) = p(t-3) - p(t-7)	-0.147*** (0.0224)	-0.211*** (0.0225)	-0.0816*** (0.0205)	-0.109*** (0.0218)	0.0989*** (0.0234)
Zipcode Fixed Effects	YES	YES	YES	YES	YES
Quarter Fixed Effects	NO	NO	NO	NO	YES
Observations	2,009	2,009	2,009	2,009	2,009
R-squared	0.817	0.828	0.85	0.851	0.952

Notes: Equation is estimated using OLS. The sample is a balanced panel consisting of 41 zipcodes in Fairfax County, VA, at a quarterly frequency from 1997-2010. Four time periods are lost in estimation due to seasonal differencing and three more due to lag length. "Zipcode" fixed effects includes a dummy variable for each zip code. "Quarter" fixed effects includes a dummy variable for each time period. ***, **, and * asterisk indicates significance at 1%, 5%, and 10% levels, respectively.

Table 1b
The Netherlands: Panel ADL Estimates
Dependent Variable: $D(t) = [\log \text{ of house prices } p(t) - \log \text{ of house prices } p(t-4)]$

Variable	[1]	[2]	Equation [3]	[4]	[5]
Bargaining power: θ (t-1)		0.0546*** (0.0079)		0.0473*** (0.00753)	0.0182** (0.00777)
Sale probability: ω (t-1)			0.0921*** (0.00506)	0.0903*** (0.00504)	0.0470*** (0.00803)
D(t-1) = p(t-1) - p(t-5)	0.9302*** (0.0173)	0.9089*** (0.0174)	0.835*** (0.0173)	0.819*** (0.0174)	0.583*** (0.0177)
D(t-2) = p(t-2) - p(t-6)	0.1551*** (0.0233)	0.1452*** (0.0232)	0.150*** (0.0222)	0.141*** (0.0221)	0.177*** (0.0200)
D(t-3) = p(t-3) - p(t-7)	-0.1866*** (0.0172)	-0.2002*** (0.0172)	-0.175*** (0.0164)	-0.187*** (0.0164)	-0.107*** (0.0173)
Region Fixed Effects	YES	YES	YES	YES	YES
Quarter Fixed Effects	NO	NO	NO	NO	YES
Observations	3204	3204	3,204	3,204	3,204
R-squared	0.839	0.842	0.855	0.856	0.897

Notes: Equation is estimated using OLS. The sample is a balanced panel consisting of 36 regions in the Netherlands, at a quarterly frequency from 1988-2010. Four time periods are lost in estimation due to seasonal differencing and three more due to lag length. "Region" fixed effects includes a dummy variable for each region. "Quarter" fixed effects includes a dummy variable for each time period. ***, **, and * asterisk indicates significance at 1%, 5%, and 10% levels, respectively.

Table 2
ADL Estimates Summary

A) Fairfax County - Parameter Signs across 41 Separate Regional Models

<u>Model 1:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Bargaining power: θ (t-1)	15	26	0	0

<u>Model 2:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Sale probability: ω (t-1)	39	2	0	0

<u>Model 3:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Bargaining power: θ (t-1)	1	34	6	0
Sale probability: ω (t-1)	36	5	0	0

F-test (5%)	<u>Reject</u>	<u>Fail to Reject</u>
($\theta=0$ and $\omega=0$)	36	5

B) The Netherlands - Parameter Signs across 36 Separate Regional Models

<u>Model 1:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Bargaining power: θ (t-1)	10	17	9	0

<u>Model 2:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Sale probability: ω (t-1)	33	3	0	0

<u>Model 3:</u>	<u>Positive, p<0.05</u>	<u>Positive, p>0.05</u>	<u>Negative, p>0.05</u>	<u>Negative, p<0.05</u>
Bargaining power: θ (t-1)	17	19	0	0
Sale probability: ω (t-1)	33	3	0	0

F-test (5%)	<u>Reject</u>	<u>Fail to Reject</u>
($\theta=0$ and $\omega=0$)	34	2

Notes: Table presents tabulations of estimates across 41 separate models in Fairfax and 36 in the Netherlands. Each model is estimated using OLS and data from each zipcode in Fairfax County and each region in the Netherlands. Model 1 estimates Table 1 - Equation [2] in each zipcode/region and tabulates the sign and statistical significance of the bargaining power parameter. Model 2 estimates Table 1 - Equation [3] in each zipcode/region and tabulates the sign and statistical significance of the sale probability parameter. Model 3 estimates Table 1 - Equation [4] in each zipcode/region and tabulates the sign and statistical significance of both parameters of interest. The overall sample is a balanced panel consisting of 41 zipcodes (36 regions) at a quarterly frequency from 1998 to 2010 (1988 to 2010) in Fairfax (the Netherlands).

Table 3
One-Quarter Forecast Summary

A) Fairfax County

Variables in VAR	Forecast sample	RMSE (in-sample)	RMSE (forecast)
a) $D(t) = [p(t) - p(t-4)]$	2006q1-2010q4	0.059	0.090
b) $D(t), \theta(t)$	2006q1-2010q4	0.053	0.078
c) $D(t), \omega(t)$	2006q1-2010q4	0.054	0.071
d) $D(t), \theta(t), \omega(t)$	2006q1-2010q4	0.049	0.066

B) The Netherlands

Variables in VAR	Forecast sample	RMSE (in-sample)	RMSE (forecast)
a) $D(t) = [p(t) - p(t-4)]$	2006q1-2010q4	0.0198	0.0149
b) $D(t), \theta(t)$	2006q1-2010q4	0.0181	0.0152
c) $D(t), \omega(t)$	2006q1-2010q4	0.0187	0.0129
d) $D(t), \theta(t), \omega(t)$	2006q1-2010q4	0.0172	0.0137

Notes: Table computes in-sample residuals and forecast errors statistics across 41 separate postal codes in Fairfax County and 36 regions in the Netherlands. Each model is estimated using an autoregressive vector model "VAR." The sample used to perform one-quarter forecasts is a balanced panel consisting of 41 zipcodes (36 regions) at a quarterly frequency from 1998 (1987) to 2010 in Fairfax (the Netherlands). Forecasts are recursively computed using information only up until the time of the forecast. This generates a balanced set of 820 forecasts (41 zipcodes X 20 quarters) in Fairfax and 720 forecasts (36 regions X 20 quarters) in the Netherlands. In-sample RMSE is computed using the model estimated from the initial sample period until 2005.

Table 4
VAR Forecast Error at Different Horizons

A) Fairfax County

Variables in VAR	Sample	RMSE (forecast), X Quarters Ahead							
		1	2	3	4	5	6	7	8
a) $D(t) = [p(t) - p(t-4)]$	2006q1-2010q4	0.090	0.135	0.179	0.219	0.253	0.285	0.313	0.335
b) $D(t), \theta(t)$	2006q1-2010q4	0.078	0.123	0.161	0.200	0.231	0.258	0.285	0.308
c) $D(t), \omega(t)$	2006q1-2010q4	0.071	0.111	0.150	0.193	0.232	0.267	0.299	0.324
d) $D(t), \theta(t), \omega(t)$	2006q1-2010q4	0.066	0.113	0.153	0.195	0.231	0.260	0.289	0.312
% Difference between [a] and [d]		35.2%	19.6%	17.2%	12.2%	9.6%	9.6%	8.3%	7.3%
F-value*		1.827*	1.429*	1.374*	1.259*	1.202*	1.201*	1.173*	1.151*
Theil's U		1.352	1.196	1.172	1.122	1.096	1.096	1.083	1.073

B) The Netherlands

Variables in VAR	Sample	RMSE (forecast), X Quarters Ahead							
		1	2	3	4	5	6	7	8
a) $D(t) = [p(t) - p(t-4)]$	2006q1-2010q4	0.0149	0.022	0.030	0.037	0.042	0.047	0.051	0.055
b) $D(t), \theta(t)$	2006q1-2010q4	0.0152	0.023	0.031	0.038	0.043	0.047	0.051	0.055
c) $D(t), \omega(t)$	2006q1-2010q4	0.0129	0.017	0.022	0.027	0.031	0.034	0.037	0.039
d) $D(t), \theta(t), \omega(t)$	2006q1-2010q4	0.0137	0.020	0.026	0.031	0.035	0.038	0.040	0.042
% Difference between [a] and [d]		8.6%	12.1%	17.2%	21.0%	22.4%	25.4%	28.5%	30.2%
F-value*		1.178*	1.257*	1.374*	1.464*	1.497*	1.572*	1.651*	1.694*
Theil's U		1.086	1.121	1.172	1.210	1.224	1.254	1.285	1.302

Notes: Table computes in-sample residuals and forecast errors statistics across 41 separate models in Fairfax and 36 regions in the Netherlands. Each model is estimated using the "VAR" command in STATA. The sample used to perform X-quarter forecasts is a balanced panel consisting of 41 zipcodes (36 regions) at a quarterly frequency from 1998 (1987) to 2010 in Fairfax (the Netherlands). Forecasts are recursively computed using information only up until the time of the forecast.