Double-question Survey Measures for the Analysis of Financial Bubbles and Crashes*

M. Hashem Pesaran[a†] and Ida Johnsson [b]

[a] Department of Economics & USC Dornsife INET,

University of Southern California, USA and Trinity College, Cambridge, UK

[b] Department of Economics, USC[‡]

January 2, 2017

Abstract

This paper proposes a new double-question survey whereby an individual is presented with two sets of questions; one on beliefs about current asset values and another on price expectations. A theoretical asset pricing model with heterogeneous agents is advanced and the existence of a negative relationship between price expectations and asset valuations is established, which is tested using survey results on equity, gold and house prices. Leading indicators of bubbles and crashes are proposed and their potential value is illustrated in the context of a dynamic panel regression of realized house price changes across a number of key MSAs in the US.

JEL Classifications: C83, D84, G12, G14.
Keywords: Price expectations, bubbles and crashes, house prices, belief valuations.

*Earlier versions of this paper have been presented at the Reserve Bank of India, Queen's University, Canada, and at York University, England. The survey questions have been designed jointly with Jeff Dominitz and Charles Manski. We are grateful to Arie Kapteyn (now at USC but previously at RAND) for his generous support of this project, and to Julie Newell, Angela Hung, and Tania Gutsche for overseeing the conduct of the surveys at RAND American Life Panel. Qiankun Zhou helped with carrying out some of the panel data regressions and Jorge Tarrasó helped with the initial analysis of the survey data. We have also received helpful comments from Ron Smith.

†Corresponding author. Email address: pesaran@usc.edu

‡Ida Johnsson was a Graduate Fellow of USC Dornsife INET while working on this project.

Conflict of Interest Disclosure

M. Hashem Pesaran: "I acknowledge research funding from the Institute for New Economic Thinking and the support of RAND with carrying out the surveys and providing us with the survey data. We are grateful to Arie Kapteyn (now at USC but previously at RAND) for his generous support of this project, and to Julie Newell, Angela Hung, and Tania Gutsche for overseeing the conduct of the surveys at RAND American Life Panel."

Ida Johnsson: "I acknowledge the Institute for New Economic Thinking for suppoting my Graduate Fellowship at USC Dornsife INET during the academic years 2014-2015 and 2015-2016. I also acknowledge the support of RAND with carrying out the surveys and providing the survey data. We are grateful to Arie Kapteyn (now at USC but previously at RAND) for his generous support of this project, and to Julie Newell, Angela Hung, and Tania Gutsche for overseeing the conduct of the surveys at RAND American Life Panel."

1 Introduction

Expectations formation is an integral part of the decision making process, yet little is known about the way individuals actually form expectations. At the theoretical level and in the context of representative agent models, the rational expectations hypothesis (REH) has gained general acceptance as the dominant model of expectations formation. But in reality markets are populated with agents that differ in a priori beliefs, information, knowledge, cognitive and processing abilities, and there is no reason to believe that such heterogeneities will be eliminated by market interactions alone.

It is true that market transactions do convey price information and reveal knowledge that could lead to expectations that are less heterogeneous as compared to heterogeneity of beliefs prior to transactions, but, as has been noted by Grossman and Stiglitz (1980), the price revelation cannot be perfect and heterogeneity is likely to be a prevalent feature of expectations across individuals. Allowing for heterogeneity of expectations is particularly important for a better understanding of bubble and crashes in asset prices. This is apparent in the theoretical literature on price bubbles where most recent contributions consider different types of traders, variously refereed to as "fundamental" and "noise" traders, or "behavioral" traders in the context of multi-type agent models. See, for example, Daniel et al. (1998), Hirshleifer (2001), Odean (1998), Thaler (1991), Shiller (2000), Shleifer (2000), and Abreu and Brunnermeier (2003). Hommes (2006) provides a survey of heterogeneous agent models in economics and finance. This is in contrast to the earlier literature, pioneered by Flood and Garber (1980), which focused on tests of transversality conditions in representative agent models. It has proved difficult to develop tests of bubbles/crashes based on representative agent models, as was recognized early on by Blanchard (1979), who concluded that "*...Detecting their [bubbles] presence or rejecting their existence is likely to prove very hard.*"

There is also a large econometrics literature on tests of asset price bubbles based on long historical time series of asset returns.[1] But the outcomes of such tests are generally inconclusive. For example, Gürkaynak (2008) after surveying a large number of studies concludes that "We are still unable to distinguish bubbles from time-varying or regime

[1]There are a few empirical studies that use panel data regressions, but such studies face the additional challenge of allowing for bubbles at different times in different markets and possible bubble spill-overs across markets.

switching fundamentals, while many small sample econometrics problems of bubble tests remain unresolved."

Recent recursive time series tests proposed in a series of papers by Phillips and Yu provide more powerful tests, but these tests are purely statistical in nature and do not allow us to infer if structural breaks detected in the time series processes of asset prices are evidence of bubbles or are due to breaks in the underlying (unobserved) fundamentals. See Phillips et al. (2011) and Phillips et al. (2015). Also see Homm and Breitung (2012).

Analysis of aggregate time series observations can provide historical information about price reversals and some of their proximate causes. But it is unlikely that such aggregate time series observations on their own could provide timely evidence of building up of bubbles and their subsequent collapse. In this paper we consider survey data on individual expectations, and exploit the considerable degree of heterogeneity of expectations documented in the literature. For example, Ito (1990) considers expectations of foreign exchange rates in Japan, and finds that exporters tend to anticipate a yen depreciation while importers anticipate an appreciation, a kind of 'wishful thinking'. Dominitz and Manski (2011) consider heterogeneity of equity price expectations using the Michigan Surveys. They find that young people tend to be more optimistic than old people about the stock market, that men are more optimistic than women, and that optimism increases with education. Branch (2004) finds that households in Michigan Surveys respond dynamically and heterogeneously when forming their expectations, with different individuals ending up using different forecasting models depending on their particular circumstances. Similar patterns of expectations heterogeneity are documented for house prices. See, for example, Case and Shiller (1988), Case and Shiller (2003), Case et al. (2012), Niu and Van Soest (2014), Kuchler and Zafar (2015), and Bover (2015). Case and Shiller find that home buyers who have experienced larger house price increases also tend to have higher expectations of future house prices.[2]

However, all surveys of price expectations focus on individual expectations of future price movements either qualitatively (whether the prices are expected to rise, fall or stay the same) or quantitatively in the form of predictive densities. The outcomes of such surveys are used in disaggregated or aggregated forms in tests of rationality of expectations and for forecasting of aggregate trends. Typically, such survey questions are not placed in particular decision contexts. However, for the analysis of many economic problems more information about

[2] A review of the literature on survey expectations can be found in Pesaran and Weale (2006).

the nature of individual beliefs and expectations is required. This is particularly the case when individual decisions depend not only on their own expectations of future outcomes, but also on their beliefs about the expectations of other market participants. But elicitation of individual expectations of others can be quite difficult. It is also likely to be unreliable since the reference group might not be known and could be changeable over time.

In this paper we consider an alternative strategy where an individual's price expectation is related to his/her subjectively held belief about the current level of prices. An individual respondent is presented with two sets of questions, one that asks about the individual's belief regarding valuations (whether the prevailing asset price is "fairly valued"), and another regarding the individual's expectations of the future price of that asset.[3] Responses to these two questions are then used to measure the extent to which prices are likely to move towards or away from the subjectively perceived fundamental values. These questions do not require that the notation of a fundamental value is commonly understood or agreed upon.

In this paper we report the results of such double-question surveys for gold, equity and house prices conducted with US households using RAND American Life Panel (ALP).[4] The ALP covers over 6,000 members with ages 18 and over, and is nationally representative, drawing from respondents recruited from several sources, including University of Michigan Phone-Panel and Internet-Panel Cohorts, and National Survey Project Cohort. We started with two pilot surveys, and introduced the double-question surveys as a new module starting in January 2012 and ended January 2013 (13 waves altogether). The number of survey participants ranged form a low of 4,477 in January 2012, to a high of 5911 in January 2013. All respondents provided demographic information, but were not compelled to respond to our questions. Nevertheless, as it turned out the response rate was around 72%, and we ended up with a panel of around 4,000 individuals who completed our survey questions over the period January 2012 to January 2013. This is a very high response rate as compared to other surveys of house prices conducted in the literature. For example, the average response rate of the homebuyers surveys conducted by Case and Shiller was around 22.7% over the

[3]The double-question surveys proposed in this paper are to be distinguished from other double-questions considered in the survey literature, such as the "double-barreled" questions that ask a respondent two questions but require one answer, and questions with anchoring vignettes, introduced by King et al. (2004), which are aimed at enhancing cross-respondent comparability of survey measures.

[4]For details of ALP see http://www.rand.org/pubs/corporate_pubs/CP508-2015-05.html. The survey questions have been designed jointly with Jeff Dominitz (Resolution Economics) and Charles Manski (Northwestern University).

years 1988, and 2003-2012. See Table 1 in Case et al. (2012).

The survey responses provide information on individuals' price expectations as well as their valuation beliefs. It is the two questions together that allow us to construct bubble and crash indicators. To our knowledge this has not been done before. The paper also makes a theoretical contribution to the literature on asset pricing with heterogeneous agents. Under certain conditions on how individuals form expectations of others in the market place, it shows that individual expectations of price changes are negatively related to their market valuation. In the absence of price bubbles/crashes, individuals who believe market prices are too high tend to have lower price expectations, whilst those who believe market prices are too low tend to have higher price expectations. However, such an error-correcting process need not hold at times of bubbles (or crashes) when individuals could believe the prices to be too high (low), and yet expect higher (lower) prices. This pattern of expectations formation is in line with theories of speculative behavior and bubbles and crashes that argue that rational traders understand that market prices might be over-valued, but continue to expect higher prices as they believe they can ride the bubble and exit just before the crash. See, for example, Abreu and Brunnermeier (2003).

The importance of heterogeneity for speculative behavior and over-valuation has been emphasized by Miller (1977). Miller was the first to show that in markets with heterogeneous agents and short-sales constraints, security prices are likely to be over-valued, since short-sales restrictions deter the pessimists from trading without a commensurate effect on the optimists. The quantitative important of this effect is investigated by Chen et al. (2002). Miller's result is obtained in a static framework, but similar outcomes are also obtained in a dynamic setting. Harrison and Kreps (1978) show that in the presence of short-sales restrictions, and when agents differ in their beliefs about the probability distributions of dividend streams, then over-valuation can arise since agents believe that in the future they will find a buyer willing to pay more than their asset's current worth. In a related paper, Scheinkman and Xiong (2003) argue that such speculative behavior can generate important bubble components even for small differences in beliefs.

These and other theoretical models of asset price over-valuation in the literature provide important insights into interactions of trader heterogeneity and other market features such as short-sales constraints. However, they are silent on the way over-valuation (or under-valuation) can affect price expectations. In this paper we consider a multi-period asset pricing model with heterogeneous traders, and show that the model has a unique bubble-free solution when traders are anonymous and individual traders base their expectations of others only on publicly available information. More importantly, for the analysis of the double-question surveys, we show that individual traders' expected price changes are related to their asset valuation, as measured by the gap between market prices and traders' own valuation. This relationship is shown to be error correcting in expectations formation, with traders who believe the market to be over-valued (under-valued) expecting prices to fall (rise). This result holds for expectations formed for longer horizons, with the weight attached to the asset valuation variable declining with the horizon. By implication, it also follows that the error correcting mechanism could become perverse if cross-agent expectations are likely to lead to indeterminate outcomes, possibly resulting in the build-up of forces for bubbles or crashes. In such situations, it is possible for traders to believe the market is over-valued (under-valued), and yet continue to expect prices to rise (fall).

We provide estimates of the relationship between expected price changes and a valuation indicator using an unbalanced panel of responses from the double-question surveys. We find statistically significant relationships between expected price changes (at one, three and twelve months ahead) and asset valuations (under or over) for all the three asset classes. But these relationships are error correcting (in the sense discussed above) for equity price expectations at longer horizons and for house price expectations at all three horizons being considered. Gold price expectations do not seem to be equilibrating. The effects of demographic factors, such as sex, age, education, ethnicity, and income are also investigated. It is shown that for house price expectations such demographic factors cease to be statistically significant once we condition on the respondents' location and their asset valuation indicator.

Finally, using the double-question survey responses we propose crash and bubble indicators for use as early warning signals of bubbles and crashes in the economy as a whole or in a particular region. There is also the issue of how to evaluate the usefulness of such indicators. One approach would be to investigate their contribution in modeling and forecasting realized price changes in a given region or nationally. A pure time series approach would require sufficiently long time series data and is not possible in the case of the present survey (which covers a very short time period). But it is possible to exploit the panel dimension of our data and see if crash and bubble indicators can significantly contribute to the explanation of *realized* house price changes across different metropolitan statistical areas (MSAs). To this end we begin with a dynamic fixed effects panel data model in monthly realized house price changes and then add expected house price changes and crash and bubble indicators at different horizons to see if such survey based indicators can help in cross-sectional explanation of realized house price changes. We employ dynamic panel data models with fixed and time effects and include MSA-specific crash and bubble indicators together with similar indicators constructed for the neighboring MSAs. We find such indicators to have significant explanatory power for realized house price changes over and above past price changes. All estimated coefficients have the correct signs, predicting expected price changes to rise with bubble indicators and to fall with the crash indicators.

The remainder of the paper is organized as follows: Section 2 sets out the theoretical asset pricing model with heterogeneous agents and derives the relationship between individual expected price changes and their asset valuations at different horizons. Section 3 describes the survey design, provides summary statistics of survey responses, and presents some preliminary data analyses. Section 4 gives the panel regressions of respondents' expected price changes on their valuation indicator, and discusses the effects of location, socio-demographic and other factors on the expectations formation process. Section 5 introduces the bubble and crash leading indicators. Section 6 investigates the importance of such leading indicators for the analysis of realized house price changes across MSAs. Section 7 ends with some concluding remarks. The exact survey questions and the filtering rules used to clean the survey data for panel regression analyses are given in the Appendix. Additional results and descriptions are provided in a Supplement which is available from the authors on request.

2 Asset pricing with heterogeneous agents

Suppose there are $n > 2$ traders with information sets $\Omega_{it} = \Phi_{it} \cup \Psi_t$, $i = 1, 2, ..., n$, respectively, where Φ_{it} is the information set specific to trader i and Ψ_t is the publicly available information set. Each trader decides on how many units, q_{it}, of a particular asset (or portfolio of assets) to hold by maximizing $E_i \left[U \left(W_{t+1,i} \right) | \Omega_{it} \right]$, where $U \left(W_{t+1,i} \right)$ represents the constant absolute risk aversion utility function

$$U \left(W_{t+1,i} \right) = 1 - \exp \left[-\gamma_i W_{t+1,i} \right],$$

where $W_{t+1,i}$ is the end of the period net worth of trader i, and γ_i is the coefficient of absolute risk aversion of the i^{th} trader, and $E_i\left(\cdot\left|\Omega_{it}\right.\right)$ is the expectations operator for trader's i conditional on his/her information set, Ω_{it}. Under this set up and assuming normally distributed asset returns and no transaction costs, it is easily established that asset demand for trader i is given by

$$P_t q_{it}^d = \frac{E_i\left(R_{t+1}\left|\Omega_{it}\right.\right) - r_t}{\gamma_i Var_i\left(R_{t+1}\left|\Omega_{it}\right.\right)},$$

where $R_{t+1} = \left(P_{t+1} - P_t + D_{t+1}\right)/P_t$, is the rate of return on holding the asset over the period t to $t+1$, P_t is the asset price at t, D_{t+1} is the dividend paid on holding the asset over period t to $t+1$, r_t is the risk free rate of return, and $Var_i\left(R_{t+1}\left|\Omega_{it}\right.\right)$ is the i^{th} trader's conditional variance of asset returns. Assuming no new shares are issued, the market clearing condition is given by $\sum_{i=1}^{n} q_{it}^d = 0$, and we have[5]

$$P_t = \left(\frac{1}{1+r_t}\right)\left[\sum_{i=1}^{n} w_{it} E_i\left(P_{t+1}\left|\Omega_{it}\right.\right) + \sum_{i=1}^{n} w_{it} E_i\left(D_{t+1}\left|\Omega_{it}\right.\right)\right], \tag{1}$$

where

$$w_{it} = \frac{\left[\gamma_i Var_i\left(R_{t+1}\left|\Omega_{it}\right.\right)\right]^{-1}}{\sum_{j=1}^{n}\left[\gamma_j Var_j\left(R_{t+1}\left|\Omega_{jt}\right.\right)\right]^{-1}}. \tag{2}$$

Equation (1) is a generalization of the standard asset pricing model and allows for the possible effects of information heterogeneity across traders on the determination of asset prices. The weights w_{it} satisfy the adding up condition, $\sum_{i=1}^{N} w_{it} = 1$, and capture the relative importance of the traders in the market.

It is well known that the solution of the above asset pricing equation is subject to the "infinite regress" problem discussed originally by Phelps (1983), Townsend (1983) and Pesaran (1987) Ch. 4. The problem arises since each trader, i, has to form expectations of the

[5]This assumption can be relaxed and replaced by $\sum_{i=1}^{n} q_{it}^d = Q$, where Q is the net addition to the supply of shares. In this case, our results hold if it is assumed that $Q/n \to 0$ as $n \to \infty$.

average future price and dividend expectations of other traders, namely

$$E_i\left[\sum_{s=1}^{n}(1+r_t)^{-1}w_{st}E_s\left(P_{t+1}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right], \text{ and}$$
$$E_i\left[\sum_{s=1}^{n}(1+r_t)^{-1}w_{st}E_s\left(D_{t+1}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right]$$

for all $s \neq i$, which in turn involves working out

$$E_i\left[\sum_{s=1}^{n}(1+r_{t+1})^{-1}w_{s,t+1}E\left(P_{t+2}\left|\Omega_{s,t+1}\right.\right)\left|\Omega_{it}\right.\right], \text{ and}$$
$$\left[\sum_{s=1}^{n}\left(1+r_{t+1}\right)^{-1}w_{s,t+1}E\left(D_{t+2}\left|\Omega_{s,t+1}\right.\right)\left|\Omega_{it}\right.\right],$$

and so on. In effect each trader needs to form expectations of other traders' price and dividend expectations for all future dates, which is a multi-period version of Keynes' well known beauty contest. In general, the solution is indeterminate even if we impose transversality conditions on all traders, individually. There are many possible solutions depending on how individual traders form expectations about the price expectations of others in the market. In what follows, to resolve the infinite regress problem and obtain an analytical baseline relationship between expected price changes and the valuation indicator, we consider a set of simplifying assumptions that allow for heterogeneity but lead to a unique bubble-free market solution. In this way we are able to model the cross section heterogeneity of expectations in an equilibrium context.

Assumption 1 *(Risk free rate) It is common knowledge that the risk free rate, r_t, is time-invariant, namely $r_t = r$.*

Assumption 2 *(Volatilities) It is common knowledge that $Var\left(R_{t+1}\left|\Omega_{it}\right.\right) = \sigma_i^2$ for all t, and $0 < c < \gamma_i\sigma_i^2 < C < \infty$, for some strictly positive constants, $c < C$.*

Assumption 3 *(Network anonymity) The traders $i = 1, 2, ..., n$ belong to an anonymous network and each trader i^{th} expectations of other traders' price expectations are given by*

$$E_i\left[E_j\left(P_{t+h}\left|\Omega_{j,t+h-1}\right.\right)\left|\Omega_{it}\right.\right] = E_i\left(P_{t+h}\left|\Omega_{it}\right.\right) + \xi_{it}^{(h)}P_t, \tag{3}$$

for all i and $j = 1, 2, ..., n$, and $h = 1, 2, ...,$ where $\xi_{it}^{(h)}$ is the idiosyncratic part of trader i^{th} expectations of trader j^{th} expectation at horizon h, and satisfy the following

$$\begin{aligned} E_i\left(\xi_{jt}^{(h)}\,|\Omega_{it}\right) &= \xi_{it}^{(h)}, \text{ for } j = i \\ &= 0, \text{ for } j \neq i. \end{aligned} \tag{4}$$

Remark 1 *The anonymity assumption ensures that trader's i^{th} expectations of trader j^{th} expectations does not depend on j.*

Assumption 4 *(Dividend processes) Traders commonly believe that the dividend process, $\{D_t\}$, follows a geometric random walk, but differ in their beliefs about the drift and volatility of the dividend process. Specifically, trader i^{th} dividend process is given by model M_i*

$$M_i : D_t = D_{t-1}\exp(\mu_i + \sigma_i\varepsilon_t), \text{ for } i = 1, 2, ..., n, \tag{5}$$

where ε_t is $i.i.d.N(0, 1)$. The true dividend process is given by

$$DGP : D_t = D_{t-1}\exp(\mu + \sigma\varepsilon_t), \tag{6}$$

Remark 2 *Conditional expectations taken under model M_i and under the DGP will be denoted by $E_i(\cdot\,|)$ and $E(\cdot\,|)$, respectively.*

Assumption 5 *(Market pooling condition) Market expectations of individual trader's price expectations are given by*

$$E\left[E_i\left(P_{t+1}\,|\Psi_t\right)|\Psi_t\right] = E\left(P_{t+1}\,|\Psi_t\right), \tag{7}$$

the transversality condition $lim_{H\to\infty}(1 + r)^{-H}E\left(P_{t+H}\,|\Psi_t\right) = 0$ holds, and $\exp(g) < 1 + r$, where $g = \mu + (1/2)\sigma^2$, with μ and σ^2 defined by (6).

Remark 3 *Assumption 5 ensures the existent of a representative agent model associated with the underlying multi-agent set up.*

To allow for market pooling of traders' disparate beliefs regarding the dividend growth process, we introduce the following assumption:

Assumption 6 *(Distribution of trader disparities) Trader-specific belief regarding his/her steady state growth rate of dividends, g_i, defined by (9), are distributed independently across i as $N(g, \omega_g^2)$.*

Under Assumptions 1 and 2, the price equation (1) simplifies to

$$P_t = \left(\frac{1}{1+r}\right)\left[\sum_{s=1}^{n} w_s E_s\left(P_{t+1}\left|\Omega_{st}\right.\right) + \sum_{s=1}^{n} w_s E_s\left(D_{t+1}\left|\Omega_{st}\right.\right)\right].$$

Also, under Assumption 4 it is easily seen that

$$E_s\left(D_{t+h}\left|\Omega_{st}\right.\right) = D_t \exp(hg_s), \tag{8}$$

where

$$g_s = \mu_s + (1/2)\sigma_s^2. \tag{9}$$

Hence

$$P_t = \left(\frac{1}{1+r}\right)\sum_{s=1}^{n} w_s E_s\left(P_{t+1}\left|\Omega_{st}\right.\right) + \theta_n D_t, \tag{10}$$

where

$$\theta_n = \frac{\sum_{s=1}^{n} w_s \exp(g_s)}{1+r}. \tag{11}$$

Now suppose that the asset pricing equation (10) is common knowledge, and is therefore used by all traders to form their price expectations and asset price valuations. In cases where expectations are homogeneous across all traders or when differences in expectations are common knowledge then applying the conditional expectations operator for the i^{th} trader, $E_i\left(\cdot\left|\Omega_{it}\right.\right)$ to both sides of (10) will yield the same result, namely P_t. However, this is not the case in the more realistic scenario where differences in expectations are not common knowledge. Clearly, for the left hand side of (10) we have $E_i\left(P_t\left|\Omega_{it}\right.\right) = P_t$ since P_t is included in Ω_{it}. But application of $E_i\left(\cdot\left|\Omega_{it}\right.\right)$ to the right hand side of (10) need not be equal to P_t since exact expressions for terms such as $E_i\left[E_s\left(P_{t+1}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right]$ are not known to trader i, and he/she has no choice but to use some form of an approximation, such as the one proposed in Assumption 3.

Accordingly, we define trader i^{th} asset valuation at time t, P_{it}^*, by applying $E_i\left(\cdot\left|\Omega_{it}\right.\right)$ to the right hand side of (10), namely

$$P_{it}^{*}=\left(\frac{1}{1+r}\right)\sum_{s=1}^{n}w_{s}E_{i}\left[E_{s}\left(P_{t+1}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right]+E_{i}\left(\theta_{n}\right)D_{t}.$$

Now under Assumption 3, and using the condition $E_{i}\left[E_{s}\left(P_{t+1}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right]=E_{i}\left(P_{t+1}\left|\Omega_{it}\right.\right)+\xi_{it}^{(1)}P_{t}$, we have

$$P_{it}^{*}=\left(\frac{1}{1+r}\right)\left[E_{i}\left(P_{t+1}\left|\Omega_{it}\right.\right)+\xi_{it}^{(1)}P_{t}\right]+E_{i}\left(\theta_{n}\right)D_{t}. \tag{12}$$

Subtracting P_t from both sides of (12) and after some re-arrangements we obtain

$$\frac{E_{i}\left(P_{t+1}\left|\Omega_{it}\right.\right)-P_{t}}{P_{t}}=-(1+r)\left(\frac{P_{t}-P_{it}^{*}}{P_{t}}\right)+\left[r-E_{i}\left(\theta_{n}\right)\left(\frac{D_{t}}{P_{t}}\right)\right]-\xi_{it}^{(1)},$$

which we write as

$$\pi_{i,t+1}^{e}=-(1+r)V_{it}+\left[r-E_{i}\left(\theta_{n}\right)\left(\frac{D_{t}}{P_{t}}\right)\right]-\xi_{it}^{(1)}, \tag{13}$$

where

$$\pi_{i,t+1}^{e}=\frac{E_{i}\left(P_{t+1}\left|\Omega_{it}\right.\right)-P_{t}}{P_{t}},\ V_{it}=\frac{P_{t}-P_{it}^{*}}{P_{t}}. \tag{14}$$

Equation (13) relates trader i^{th} expected rate of price change to his/her over- or under-valuation of the asset, as measured by V_{it}. Note that θ_n is not known to trader i and $E_i\left(\theta_n\right)$ represents trader i^{th} expectations of θ_n.

In equilibrium the realized price dividend-ratio, P_t/D_t, is determined by taking expectations of the asset pricing equation (10) conditional on the publicly available, Ψ_t, across all traders. Specifically, we have

$$\begin{aligned}E\left(P_{t}\left|\Psi_{t}\right.\right) &= P_{t}=\left(\frac{1}{1+r}\right)\sum_{i=1}^{n}w_{i}E\left[E_{i}\left(P_{t+1}\left|\Omega_{it}\right.\right)\left|\Psi_{t}\right.\right]+E\left(\theta_{n}\right)D_{t},\\ &= \left(\frac{1}{1+r}\right)\sum_{i=1}^{n}w_{i}E\left[E_{i}\left(P_{t+1}\left|\Psi_{t}\right.\right)\left|\Psi_{t}\right.\right]+E\left(\theta_{n}\right)D_{t}.\end{aligned}$$

Further by Assumption 5 we have (recall that $\Sigma_{i=1}^{n} w_i = 1$)

$$P_t = \left(\frac{1}{1+r}\right) E\left(P_{t+1}\left|\Psi_t\right.\right) + E\left(\theta_n\right) D_t.$$

This is a standard asset pricing model for a representative risk neutral agent with the dividend process given by (6). Under standard transversality condition applied to P_t, it has the following unique solution:

$$P_t = E\left(\theta_n\right) \sum_{j=0}^{\infty} \left(\frac{1}{1+r}\right)^j E\left(D_{t+j}\left|\Psi_t\right.\right),$$

which in view of (6) yields (recall that $\exp(g) < 1+r$)

$$P_t/D_t = \frac{(1+r)E\left(\theta_n\right)}{1+r-e^g} = \frac{\sum_{s=1}^{n} w_s E\left[\exp(g_s)\right]}{1+r-e^g}. \tag{15}$$

Using this result in (13) now gives the following relationship between expectations and valuations

$$\pi_{i,t+1}^e = \alpha_i - (1+r)V_{it} + u_{it}, \tag{16}$$

where, as before, $\pi_{i,t+1}^e = E_i\left(\pi_{t+1}|\Omega_{it}\right)$, $\pi_{t+1} = (P_{t+1} - P_t)/P_t$, $V_{it} = \left(P_t - P_{it}^*\right)/P_t$, and

$$\alpha_i = r - \frac{E_i\left(\theta_n\right)\left(1+r-e^g\right)}{E\left(\theta_n\right)}, \text{ and } u_{it} = -\xi_{it}^{(1)}. \tag{17}$$

It is easily seen that in the homogenous information case where, $\Omega_{it} = \Psi_t$, and $g_i = g$, then we also have $P_{it}^* = P_t$, and $E_i\left(\theta_n\right) = E\left(\theta_n\right)/D_t$, for all i. Furthermore, (16) reduces to $\pi_{i,t+1}^e = e^g - 1$, for all i.

Another interesting feature of the above solution is that the equilibrium price-dividend ratio under heterogeneous information is *strictly larger* than the ratio obtained under homogeneity. This follows from (15) and by noting that under homogeneity the price-dividend ratio is given by $e^g/\left(1+r-e^g\right)$, whilst under heterogeneous g_i it is given by $e^{g+0.5\omega_g^2}/\left(1+r-e^g\right)$, with $1+r > e^g$ and $\omega_g^2 > 0$. This finding mirrors the over-valuation results due to Miller (1977) and Harrison and Kreps (1978), discussed in the Introduction, but holds more generally even in the absence of short-sales constraints. The extent of over-valuation under heterogeneity depends on the dispersion of opinion across traders about g_i, the rate of growth of the dividends.

2.1 Higher-order ahead expectations and valuations

The error-correction specification (16) can be generalized to price expectations for higher-order horizons. Advancing both sides of equation (10) one period ahead we first note that,

$$P_{t+1} = \left(\frac{1}{1+r}\right) \sum_{s=1}^{n} w_s E_s \left(P_{t+2} \left| \Omega_{s,t+1} \right.\right) + \theta_n D_{t+1},$$

and applying the conditional expectations operator, $E_i \left(\cdot \left| \Omega_{it} \right.\right)$ we have

$$E_i \left(P_{t+1} \left| \Omega_{it} \right.\right) = \left(\frac{1}{1+r}\right) \sum_{s=1}^{n} w_s E_i \left\{ \left[E_s \left(P_{t+2} \left| \Omega_{s,t+1} \right.\right)\right] \left| \Omega_{it} \right.\right\} + E_i \left(\theta_n\right) D_t e^{g_i}.$$

But by (3), $E_i \left[E_s \left(P_{t+2} \left| \Omega_{s,t+1} \right.\right) \left| \Omega_{it} \right.\right] = E_i \left(P_{t+2} \left| \Omega_{it} \right.\right) + \xi_{it}^{(2)} P_t$, and we have

$$E_i \left(P_{t+1} \left| \Omega_{it} \right.\right) = \left(\frac{1}{1+r}\right) \left[E_i \left(P_{t+2} \left| \Omega_{it} \right.\right) + \xi_{it}^{(2)} P_t \right] + E_i \left(\theta_n\right) D_t e^{g_i}.$$

Substituting this result in (10)

$$P_t = \left(\frac{1}{1+r}\right) \sum_{i=1}^{n} w_i \left\{ \left(\frac{1}{1+r}\right) \left[E_i \left(P_{t+2} \left| \Omega_{it} \right.\right) + \xi_{it}^{(2)} P_t \right] + E_i \left(\theta_n\right) D_t e^{g_i} \right\} + \theta_n D_t,$$

and after some simplification we have

$$P_t = \left(\frac{1}{1+r}\right)^2 \sum_{s=1}^{n} w_s E_s \left(P_{t+2} \left| \Omega_{st} \right.\right) + \left(\frac{1}{1+r}\right)^2 \left(\sum_{s=1}^{n} w_s \xi_{st}^{(2)}\right) P_t + \phi_n D_t, \tag{18}$$

where

$$\phi_n = \left(\frac{1}{1+r}\right) \left(\sum_{s=1}^{n} w_s E_s \left(\theta_n\right) e^{g_s}\right) + \theta_n. \tag{19}$$

As before P^*_{it} is defined by applying the expectations operator $E_i\left(P_t\left|\Omega_{it}\right.\right)$ to the right hand side of (18), namely

$$\begin{aligned}
P^*_{it} &= \left(\frac{1}{1+r}\right)^2 \sum_{s=1}^{n} w_s E_i\left[E_s\left(P_{t+2}\left|\Omega_{st}\right.\right)\left|\Omega_{it}\right.\right] \\
&\quad + \left(\frac{1}{1+r}\right)^2 \left[\sum_{s=1}^{n} w_s E\left(\xi^{(2)}_{st}\left|\Omega_{it}\right.\right)\right] P_t + E_i\left(\phi_n\right) D_t.
\end{aligned}$$

Now using (3) and (4) in the above equation yields

$$\begin{aligned}
P^*_{it} &= \left(\frac{1}{1+r}\right)^2 \left[E_i\left(P_{t+2}\left|\Omega_{it}\right.\right) + \xi^{(2)}_{it} P_t\right] \\
&\quad + \left(\frac{1}{1+r}\right)^2 w_i \xi^{(2)}_{it} P_t + E_i\left(\phi_n\right) D_t.
\end{aligned}$$

Subtracting P_t from both sides, using (15), and after some simplifications, and obtain

$$\pi^e_{i,t+2} = \alpha^{(2)}_i - \frac{(1+r)^2}{2} V_{it} + u^{(2)}_{it},$$

where

$$\begin{aligned}
\pi^e_{i,t+2} &= E_i\left(\pi_{t+2}\left|\Omega_{it}\right.\right),\ \pi_{t+2} = \frac{P_{t+2}-P_t}{2P_t} = \frac{\Delta P_{t+2} + \Delta P_{t+1}}{2P_t}, \\
\alpha^{(2)}_i &= \frac{(1+r)^2 - 1}{2} - \frac{(1+r)\left(1+r-e^g\right) E_i\left(\phi_n\right)}{2E\left(\theta_n\right)}, \\
u^{(2)}_{it} &= -\left(\frac{1+w_i}{2}\right)\xi^{(2)}_{it}.
\end{aligned}$$

As before, under information homogeneity, $u^{(2)}_{it} = 0$, $V_{it} = 0$ and $\pi^e_{i,t+2} = \alpha^{(2)}_i = \left(e^{2g}-1\right)/2$, for all i.

In general, for a finite h we have

$$\pi^e_{i,t+h} = \alpha^{(h)}_i - \frac{(1+r)^h}{h} V_{it} + u^{(h)}_{it}, \tag{20}$$

where $\pi^e_{i,t+h} = E_i\left(\pi_{t+h}\left|\Omega_{it}\right.\right)$, $\pi_{t+h} = (P_{t+h} - P_t)/hP_t$, and $\alpha^{(h)}_i$ and $u^{(h)}_{it}$ can be obtained similarly. For the empirical analysis to follow, it is sufficient to note that the asset valuation coefficient, $(1+r)^h/h$, tends to fall with h for small values of r and so long as h is not too large.

3 Double-question surveys

To our knowledge the use of double-question surveys to elicit a respondent asset valuation along with her/his price expectations is new. Whilst there is a large and expanding literature on surveys of price expectations, there is no attempt at direct measurement of individual's valuation of asset prices. We needed to carry out a fresh survey that simultaneously included both questions on expectations and valuations. With this in mind and in collaboration with Jeff Dominitz and Charles Manski, we designed survey questions on expectations and valuations for US households, using RAND American Life Panel (ALP).[6]

The ALP has a modular form, which allowed us to combine demographic, education and income data with the results from our double-question surveys. The double-question surveys on belief and expectations added to the ALP surveys covered equity, gold, and house prices. The two questions for equity prices were as follows.

[6] We are particularly grateful to Arie Kapteyn (now at USC but previously at RAND) for his generous support of this project. The sampling frame of ALP surveys, and other details can be found from the following link http://www.rand.org/pubs/corporate_pubs/CP508-2016-04.html.

Question 1 (equity)
We have some questions about the price of publicly traded stocks. Do you believe the US stock market (as measured by S&P 500 index) to be currently:
1 Overvalued
2 Fairly valued (in the sense that the general level of stock prices is in line with what you personally regard to be fair)
3 Undervalued

Note: The S&P 500 is an index of 500 common stocks actively traded in the United States. It provides one measure of the general level of stock prices.

Question 2 (equity)
Bearing in mind your response to the previous question, suppose now that today someone were to invest 1000 dollars in a mutual fund that tracks the movement of S&P 500 very closely. That is, this "index fund" invests in shares of the companies that comprise the S&P 500 Index. What do you expect the $1000 investment in the fund to be worth
- in one month from now,
- in three months from now,
- in one year from now.

We also asked the respondents a third question regarding the chance of $1,000 investment to fall in three different ranges. Further details can be found in Appendix A.1. A similar set of questions was asked about gold prices.

The second set of questions is on house prices in the metropolitan area where the respondent is resident. Respondents were provided with the median price of a single family home in the area close to their place of residence. We used quarterly house prices disaggregated by 180 MSAs from the National Association of Realtors.[7] This turned out to be an important consideration given the heterogeneity of house prices and their trajectories across the US. Although, due to privacy considerations APL does not provide ZIP code information on respondents, we were able to match respondents to MSAs using their self-reported city and state of residence. Respondents who resided further than 500 miles away from a major metropolitan area were instead asked about the median US house price. The survey questions on house prices for respondents who resided closer than 500 miles away from a major

[7] All areas are metropolitan statistical areas (MSA) as defined by the US Office of Management and Budget though in some areas an exact match is not possible from the available data. For further details see http://www.realtor.org/topics/existing-home-sales.

metropolitan area are presented below. The exact wording of the survey questions can be found in the appendix. See Appendix A.1.

Question 1 (house prices)
We now have some questions about housing prices. The median price of a single family home in the [*fill for city nearest to R zip code*] cosmopolitan area is currently around [*converted fill for median housing price in R zip code area*] (Half of all single family homes in the area cost less than the median, and the other half cost more than the median.). Do you believe that current housing prices are:
1 just right (in the sense that housing prices are in line with what you personally regard to be fair),
2 too high,
3 too low as compared to the fair value?

Question 2 (house prices)
Bearing in mind your response to the previous question, suppose now that someone were to purchase a single family home in [*fill for city nearest to R zip code*] area for the price of [...] What do you expect the house to be worth (Please enter a numeric answer only, with no commas or punctuation)
- 1 month from now,
- 3 months from now,
- 1 year from now.

3.1 Survey waves and respondent characteristics

The American Life Panel (ALP) consists of over 6,000 panel members aged 18 and older. Participants are recruited from various sources, such as the University of Michigan phone-panel and internet-panel and cohorts, mailing experiments, phone experiments and vulnerable population cohorts. The panel is representative of the nation, and panel members are provided with equipment that allows them to respond any survey programmed by RAND. The attrition rate of ALP participants is relatively low, between 2006 and 2013 the annual attrition rates were between 6 and 13 per cent. Panel members who have answered a non-household information survey within the last year are considered active and are invited to surveys. Each survey, in addition to the specific survey questions, contains a "Demographics" module, which elicits demographic and socio-economic information about the respondent.

After conducting two pilot surveys for the Double-Question Survey module, thirteen survey waves were fielded on the third Monday of each month beginning in January 2012 and ending in January 2013. ALP members were offered the opportunity to respond to our double-question (Double-Q) surveys, but their participation was not made mandatory. Table 1 provides the number of ALP members who participated in the surveys and the fraction of those who completed the double-question surveys. The response rates were quite high and averaged around 72 per cent of the survey participants, and varied little across the 13 survey waves. We found no significant demographic differences between the respondents and non-respondents of our double-question surveys.[8]

Table 1: **Survey waves and response rates**

Waves	**Months**	**All ALP participants**	**Completed Double Q Surveys**		**Filtered Samples**	
				per cent$^{(1)}$		per cent$^{(2)}$
1	January 2012	4477	3371	75	2707	80
2	February 2012	4864	3685	75	2727	74
3	March 2012	5015	3721	74	2991	80
4	April 2012	5260	3723	71	2967	80
5	May 2012	5464	3706	68	2982	80
6	June 2012	5568	4179	75	3379	81
7	July 2012	5674	4135	73	3363	81
8	August 2012	5713	4208	74	3445	82
9	September 2012	5762	4162	72	3425	82
10	October 2012	5772	4180	72	3421	82
11	November 2012	5847	3926	67	3169	81
12	December 2012	5894	4083	69	3404	83
13	January 2013	5911	4209	71	3415	81

The surveys were fielded on the third Monday of the month
$^{(1)}$ - Respondents who completed the Double Question Surveys as a percentage of all ALP participants
$^{(2)}$ - Filtered respondents as percentage of all respondents who completed the Double Question Surveys

3.2 Filters applied to survey responses

To reduce the impact of extreme outlier responses on our analysis we applied a number of filters to the responses. We also dropped waves 1 and 2 since, as was noted above, in the case of these waves respondents residing more than 500 miles from major metropolitan areas were not provided with house price data. This shortcoming was rectified in the subsequent waves

[8] The ALP surveys allow us to obtain the demographic characteristics of all survey participants even those who did not complete our questions.

(3-11), by providing such respondents with US median house prices. For these remaining survey waves (March 2012 to January 2013), we ended up with 5,480 respondents. We applied the following truncation filters to the data. First, we dropped all respondents with missing responses to the survey questions or missing demographic characteristics. We also dropped respondents whose demographic characteristics were incomplete or contained inconsistent entries over time.[9] Finally, for all expectations horizons (one month, three months and one year) and for all asset prices (equity, gold, housing) we remove respondents from our analysis if they

1. reported an expected price equal to zero for any of the survey questions,
2. reported any expected price rises for equity or gold which were in excess of 400 per cent,
3. reported expected price rises for equity or gold for all horizons in excess of 200 per cent, or reported expected price falls of more than 90 per cent for all expectations horizons,
4. reported expected house price rises in excess of 200 per cent, or expected house price falls of less than 50 per cent for all expectation horizons.

The application of the above filters removed 18.7 per cent of the total responses, leaving us with 35,961 responses and 4,971 respondents. In Table 3 we compare the demographic characteristics of the original and filtered samples for all thirteen waves. Around 20 per cent of the responses were filtered in any given survey wave. The percentage of Black and Hispanic/Latino respondents is slightly lower in the filtered sample. Also, respondents in the filtered sample have a higher average household income and education as compared to the original unfiltered sample.

The frequency distribution of monthly participation of the respondents in the filtered sample is shown in Table 2. Just over a quarter of respondents (1,268) answered the double-question surveys for all the 11 waves (3 to 13), 50 per cent (2,453) answered 9 waves, suggesting a high degree of over-time participation of the respondents in the double-question surveys.

[9] Detailed descriptions are provided in Appendix A.2.

Table 2: **Empirical frequency distribution of participants by months**

Months	11	10	9	8	7	6	5	4	3	2	1
No.	1268	1933	2453	2779	3088	3331	3597	3860	4161	4520	4971
Per cent	25.51	38.89	49.35	55.90	62.12	67.01	72.36	77.65	83.71	90.93	100

The average and median number of months participated are 7.23 and 6, respectively. The distribution is based on respondents who remained in the sample after the truncation filter is applied.

Table 3: **Comparison of original and filtered respondent samples**

Wave	**Age**		**Income**		**Female**		**Asian**		**Black**		**Hispanic/Latino**		**Education**	
	average		*average*		*per cent*		*per cent*		*per cent*		*per cent*		*average*	
	unfiltered	filtered	unfiltered	filtered	unfiltered	filtered	unfiltered	filtered	unfiltered	filtered	unfiltered	filtered	unfiltered	filtered
3	49.05	49.77	52,887	56,683	59.62	57.81	2.13	2.31	10.76	8.29	17.39	14.61	1.33	1.39
4	49.04	49.84	51,965	56,507	59.17	56.99	2.05	2.33	11.01	8.70	18.27	14.32	1.31	1.38
5	49.09	49.86	51,285	55,962	58.96	56.84	1.78	1.98	11.10	8.72	17.84	14.52	1.30	1.37
6	48.90	49.47	51,736	56,039	58.78	56.76	1.89	2.13	11.52	8.97	18.53	15.54	1.31	1.37
7	48.70	49.33	51,518	55,240	59.51	57.57	1.86	1.96	11.84	9.43	18.58	15.73	1.31	1.37
8	48.86	49.50	51,967	55,444	59.62	57.53	1.95	2.03	11.19	9.20	18.99	15.82	1.31	1.37
9	48.99	49.66	51,423	54,983	59.35	57.93	1.78	1.87	11.57	9.69	18.50	16.00	1.30	1.36
10	49.11	49.69	51,900	55,689	59.09	56.85	1.87	1.99	11.59	9.68	19.07	15.81	1.31	1.37
11	49.02	49.90	52,003	56,105	59.20	57.43	1.94	1.99	11.19	9.02	18.11	14.52	1.31	1.38
12	49.33	49.93	51,423	54,992	59.10	57.26	1.84	2.00	11.53	9.37	19.40	16.75	1.30	1.36
13	48.78	49.44	51,659	55,565	58.98	57.19	1.93	2.17	11.70	9.22	18.50	15.26	1.31	1.38
Average	48.99	49.67	51,797	55,746	59.22	57.29	1.91	2.07	11.36	9.12	18.47	15.35	1.31	1.37

(1) - original sample of 5,480 respondents
(2) - filtered sample of 4,971 respondents
Education is equal to 0 if the respondent has no high school diploma, 1 if the respondent is a high school graduate with a diploma, some college but no degree, an associate degree in college occupational/vocational or academic program, and 2 if the respondent has a Bachelor's degree or higher.

3.3 Socio-demographic characteristics of respondents:

For the purposes of the econometric analysis, we calculate respondent-specific time averages of the variables age, income and education. A summary of selected socio-demographic characteristics of the respondent sample is presented in Table 4. Female respondents constitute 59 per cent of the sample, and are thus slightly over-represented compared to 51 per cent for the entire US population. A comparison of the age distribution, ethnicity and educational attainment of the respondents and the entire US population is presented in Figures 1 to 4. The main differences between the respondents remaining in our sample and the US population are as follows:

- The age group 50 to 70 years old constitute a higher fraction of the ALP respondent sample compared to the US population.
- Roughly 2 per cent of the respondents identify as Asian or Pacific Islanders, the corresponding number for the entire US population is 5.4 per cent.
- ALP respondents have a higher educational level than the US population.
- Households with an annual income higher than $125,000 are under-represented in the ALP respondent sample.

Figure 1: Age distribution of ALP respondents and US population

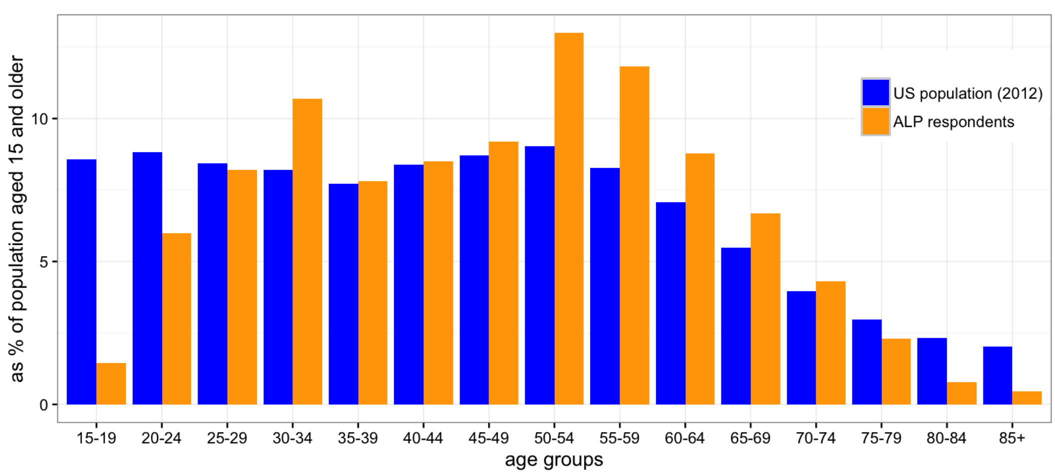


Table 4: **Summary statistics of respondent-specific time invariant characteristics**

Statistic	Mean	St. Dev.	Min	Median	Max
Age	47.80	15.50	16	49	94
Family income[1] ($)	52,470	36,627	5,000	45,000	200,000
Female (%)	0.59	0.49	0	1	1
Asian (%)	0.02	0.14	0	0	1
Black (%)	0.11	0.31	0	0	1
Hispanic/Latino (%)	0.19	0.39	0	0	1
Education Index[2]	1.33	0.57	0	1	2

All statistics are based on the sample of 4,971 respondents.

[1] - note that incomes higher than 200,000 were coded as equal to 200,000

[2] - respondent's education averaged over the time period the respondent participated in the survey, where education is equal to 0 if the respondent has no high school diploma, 1 if the respondent is a high school graduate with a diploma, some college but no degree, an associate degree in college occupational/vocational or academic program, and 2 if the respondent has a Bachelor's degree or higher.

The ALP distributions are based on the sample of 4,971 respondents.

The data on US population is obtained from the following sources:

http://www.census.gov/population/age/data/2012comp.html

https://www.census.gov/popest/data/historical/2010s/vintage _2012/national.html

Figure 2: Ethnicity of ALP respondents and US population

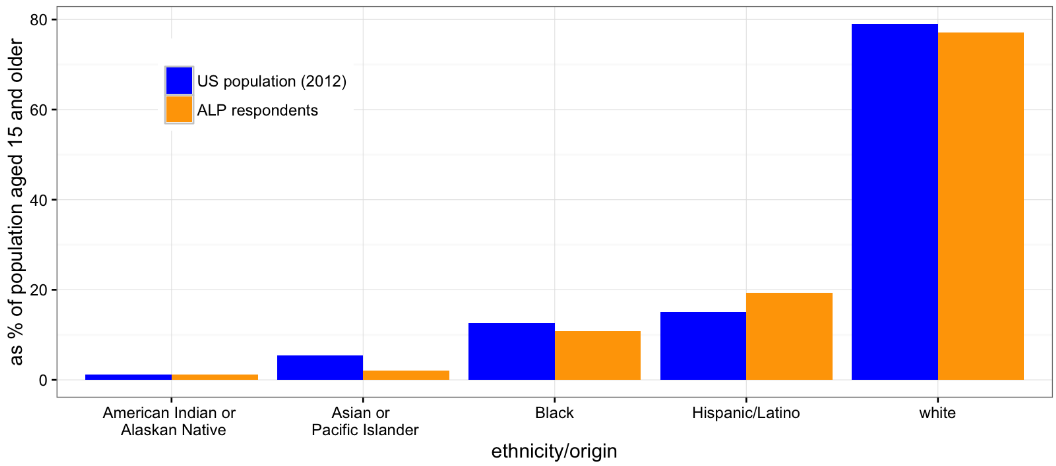
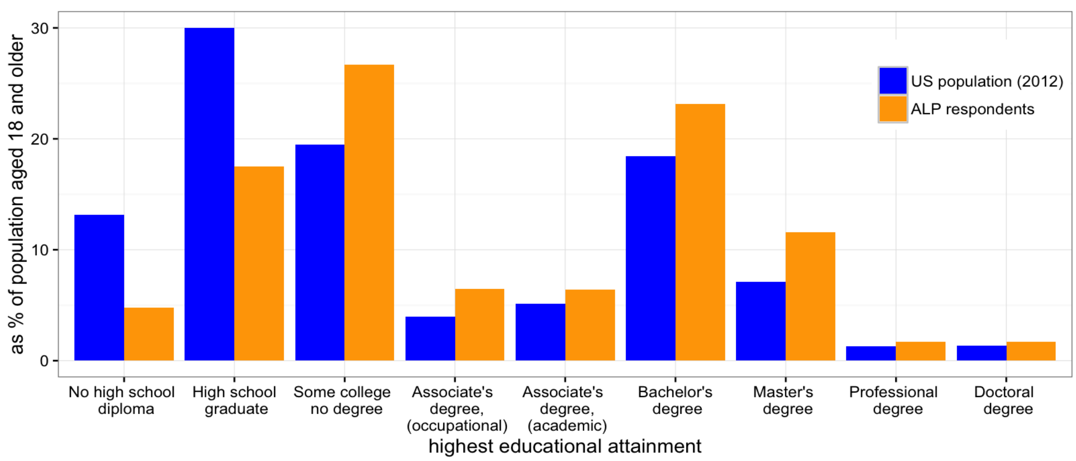


Figure 3: Educational attainment of ALP respondents and US population

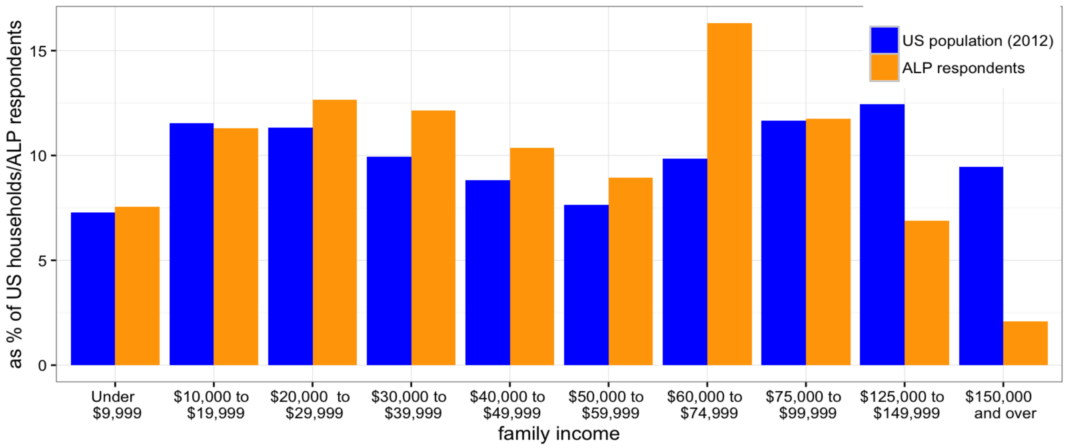


The ALP education distribution is based on 4,968 (out of 4,971) respondents who are aged 18 or older.
The ALP income distribution is based on the sample of 4,971 respondents.
The data on US population is obtained from the following sources:
http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html
http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-06.2012.html .

3.4 Geographic location of respondents

The geographic location of the respondents is shown in Figure 5, with the US population density displayed in Figure 6. Around 20 per cent of the respondents in any given survey wave

Figure 4: Income distribution of ALP respondents and US population



resided further than 500 miles away from a major metropolitan area, and were thus given the median US house price instead of the local house price in the survey section on house prices. From the sample of 4,971 respondents, we could match exactly 4,000 to a Metropolitan Statistical Area. We achieved this using the information about the respondent's city and state of residence, provided in the survey.[10]

The geographical distribution of the respondents across the eight Mainland US regions together with national figures for the US in 2012 are provided in Table 5. We first note that the geographical distribution of the respondents over time is relatively stable, which reflects the high degree of their over-time participation in the double-question surveys. The geographical distribution of the respondents also match closely the national distribution for the six out of the eight regions. The exceptions are South East and South West. Survey respondents are underrepresented in the South East region and over-represented in the South West region.

Overall, the above comparative analysis suggests that the double-question sample of respondents are fairly typical of the US population and provide a reasonable mix of individuals with different demographic and location characteristics. Furthermore, to allow for unobserved characteristics of individual respondents (such as their optimistic or pessimistic disposition) we focus primarily on the fixed effects estimates and report the full set of random effect estimates in an online supplement.

[10] As noted earlier, we did not have access to survey respondents' zip codes.

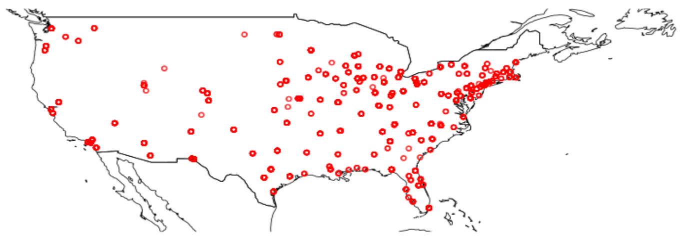


Figure 5: Respondent location

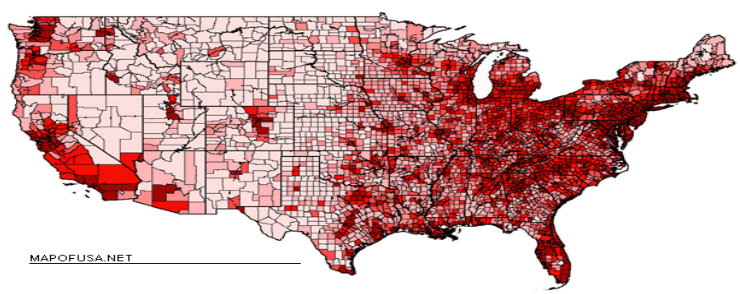


Figure 6: US population density

Table 5: **Distribution of Double Q Survey respondents and US population by region**

	Waves	Far West	Great Lakes	Mid-East	New England	Plains	Rocky Mountains	South East	South West
Double Q Respondents*	3	0.19	0.16	0.15	0.04	0.05	0.05	0.17	0.19
	4	0.18	0.17	0.15	0.04	0.05	0.05	0.17	0.19
	5	0.18	0.18	0.14	0.04	0.05	0.05	0.17	0.19
	6	0.18	0.17	0.15	0.04	0.05	0.05	0.17	0.20
	7	0.18	0.16	0.15	0.04	0.05	0.05	0.18	0.19
	8	0.18	0.16	0.14	0.04	0.05	0.05	0.18	0.19
	9	0.18	0.16	0.15	0.04	0.05	0.05	0.18	0.19
	10	0.18	0.16	0.15	0.04	0.05	0.05	0.18	0.19
	11	0.18	0.17	0.15	0.04	0.05	0.05	0.17	0.19
	12	0.18	0.16	0.15	0.04	0.05	0.05	0.18	0.20
	13	0.19	0.17	0.15	0.03	0.05	0.05	0.17	0.19
US	2012	0.17	0.15	0.15	0.05	0.07	0.04	0.25	0.12

* - filtered sample

4 Panel regressions on price expectations and valuation indicators

We are now in a position to provide empirical evidence on the importance of individual asset valuations, V_{it}, on expected prices changes, as set out in (20). Bearing in mind the survey questions, the expected rate of price changes, $\pi^e_{i,t+h|t}$, are computed by

$$\hat{\pi}^e_{i,t+h|t} = 100 \times \frac{P^e_{i,t+h|t} - 1000}{1000 \times h}, \tag{21}$$

in the case of equity and gold prices, and by

$$\hat{\pi}^e_{i,t+h|t} = 100 \times \frac{P^e_{i,t+h|t} - P^0_{it}}{P^0_{it} \times h}, \tag{22}$$

for house prices, where $P^e_{i,t+h|t}$ is the i^{th} respondent's price expectation formed at time t for h months ahead, and P^0_{it} is the house price provided to the respondent i at time t. We assume that

$$\pi^e_{i,t+h|t} = \hat{\pi}^e_{i,t+h|t} + \eta_{i,t+h}, \tag{23}$$

where $\eta_{i,t+h}$ is the error associated with the measurement of $\pi^e_{i,t+h|t}$. Using responses to the first question of the surveys we measure $sign\,(V_{it})$, by x_{it} with $x_{it} = 1$ if respondent i at time t believes the asset is over-valued ($i.e.$ $V_{it} > 0$), $x_{it} = -1$ if respondent i at time t believes the asset is under-valued ($V_{it} < 0$), and $x_{it} = 0$, otherwise. We then approximate V_{it} by $\phi_i x_{it}$, with $\phi_i > 0$, is a scalar constant. Setting $\phi_i = \phi + \zeta_i$, and using the above results in (20), we obtain

$$\hat{\pi}^e_{i,t+h|t} = \alpha^{(h)}_i + \beta^{(h)} x_{it} + \varepsilon_{i,t+h}, \tag{24}$$

where

$$\beta^{(h)} = -\frac{\phi\,(1+r)^h}{h}, \text{ and } \varepsilon_{i,t+h} = u^{(h)}_{it} - \frac{(1+r)^h}{h}\zeta_i x_{it} - \eta_{i,t+h}. \tag{25}$$

We estimate $\beta^{(h)}$ for the three asset classes assuming that ζ_i and $\eta_{i,t+h}$ are independently distributed over i and of the valuation indicator, x_{it}. These assumptions ensure that x_{it} and $\varepsilon_{i,t+h}$ are uncorrelated. We also allow for common (economy-wide) effects on individual expectations by including a time effect in (24), which gives the following fixed-effects, time-effects (FE-TE) panel regression

$$\hat{\pi}^e_{i,t+h|t} = \alpha_i^{(h)} + \beta^{(h)} x_{it} + \delta_t^{(h)} + \varepsilon_{i,t+h}. \tag{26}$$

This is a reasonably general framework that allows for random errors in measurement of expectations, random heterogeneity in the scale parameters ϕ_i, and possible time effects.

We provide estimates of $\beta^{(h)}$ for the three different asset classes, and for all the three horizons, $h = 1, 3$, and 12, separately. We use the full set of responses which yields an unbalanced panel and estimate (26) with and without time effects, allowing the individual effects, $\alpha_i^{(h)}$, to be correlated with $\varepsilon_{i,t+h}$ (and hence with its components, $\zeta_i x_{it}$, $u_{it}^{(h)}$, and $\eta_{i,t+h}$). We report FE and FE-TE estimates of $\beta^{(h)}$, together with standard errors robust to serially correlated and heteroskedastic errors in Table 6.

Table 6: **Estimates of $\beta^{(h)}$ in the panel regressions of individual expected price changes on their belief valuation indicators for different assets (equation (26))**

Dependent variable: $\hat{\pi}^e_{i,t+h|t}$

Assets	Equity		Gold		Housing	
Horizons	FE	FE-TE	FE	FE-TE	FE	FE-TE
One Month	-0.0991	-0.126	0.602***	0.581***	-0.292***	-0.303***
Ahead ($h = 1$)	(0.127)	(0.128)	(0.197)	(0.198)	(0.0643)	(0.0642)
Three Months	-0.0905	-0.0995	0.222**	0.203*	-0.106***	-0.109***
Ahead ($h = 3$)	(0.0760)	(0.0760)	(0.108)	(0.109)	(0.0273)	(0.0274)
One Year	-0.115***	-0.117***	-0.0226	-0.0316	-0.0481***	-0.0479***
Ahead ($h = 12$)	(0.0365)	(0.0364)	(0.0488)	(0.0489)	(0.0102)	(0.0102)
	$N = 35,961$, $T_{min} = 1$, $T_{p25} = 4$, $T_{p50} = 6$, $\bar{T} = 7.23$, $T_{p75} = 9$, $T_{max} = 11$					

Fixed effect (FE) estimates of $\beta^{(h)}$ in the panel regression $\hat{\pi}^e_{i,t+h|t} = \alpha_i^{(h)} + \beta^{(h)} x_{it} + u_{it}^{(h)}$ are obtained with and without time effects (FE-TE) using an unbalanced panel of 4,971 respondents over 11 months, March 2012 to January 2013.

Standard errors are in parentheses, *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively. Standard errors are robust to heteroskedasticity and residual serial correlation.

The FE estimates of $\beta^{(h)}$ for equity price expectations are statistically insignificant for $h = 1$ and 3, but become statistically significant and negative for $h = 12$. These results are in line with our theoretical findings and suggest that over the sample under consideration equity price expectations and belief valuations are consistently related. However, the same is not true of the results for gold prices, where $\beta^{(h)}$ is estimated to be positive and statistically significant for $h = 1$ and 3, and suggest that respondents might view gold prices to be over-valued and still expect gold prices to rise. Interestingly enough, even for gold prices $\beta^{(h)}$ stops being statistically significant for $h = 12$, suggesting the short term nature of the misalignment between expectations and valuations. By contrast, the estimates of $\beta^{(h)}$ for house prices are much more coherent across h and are all negative and statistically highly significant. Also, FE estimates of $\beta^{(h)}$ for house prices fall with h, as predicted by the theory. Similar conclusions are obtained if the FE-TE estimates are considered.

Although, the scaling parameter ϕ is not identified, an estimate of r, the discount rate can be obtained using any two of the estimates $\hat{\beta}^{(h_1)}$ and $\hat{\beta}^{(h_2)}$, so long as $|\hat{\beta}^{(h_1)}| > |\hat{\beta}^{(h_2)}|$. More specifically, using $\beta^{(h)} = -h^{-1}\phi\,(1+r)^h$ we have

$$\hat{r}(h_1, h_2) = \left(\frac{h_1}{h_2} \frac{\hat{\beta}^{(h_1)}}{\hat{\beta}^{(h_2)}} \right)^{\frac{1}{h_1 - h_2}} - 1. \tag{27}$$

The various estimates of r using $(h_1, h_2) = (1, 3), (3, 12)$ and $(1, 12)$, for both FE and FE-TE estimates of $\beta^{(h)}$ are summarized in Table 7. The estimates for r range between 4.0 to 6.9 percent, although given the ambiguity surrounding longer term expectations the estimates based on $\beta^{(h)}$ for $h = 1$ and $h = 3$, namely around 4 per cent, are likely to be more reliable.

Overall, the panel estimates support the predictions of the heterogenous agent model developed in Section 2, and suggest a strong relationship between respondent's housing price expectations and their valuations which is shown to be equilibrating, at least over the period under consideration. The same cannot, however, be said about the gold price expectations. This could be due to the fact that respondents are likely to have more first hand knowledge and experience about house prices as compared to international gold prices. The results for equity prices are ambiguous; there are no statistically significant relationship between equity price expectations and valuations at one month and three months horizons, which is in line

Table 7: **Alternative estimates of the discount rate, r, using FE and FE-TE estimates of $\beta^{(h)}$ for house prices**

	Based on FE estimates	Based on FE-TE estimates
$\hat{r}(1,3)$	0.044	0.039
$\hat{r}(1,12)$	0.064	0.060
$\hat{r}(3,12)$	0.069	0.065

with the prediction of a representative agent model. Nevertheless, for one year horizons asset valuations seem to play a significant role in respondent's price expectations formation process.

4.1 Effects of individual-specific characteristics on price expectations

So far we have focused on the effects of valuations on price expectations, and by using a fixed effects panel data set up, we have shown our results to be robust to individual-specific heterogeneity. But it is also of interest to investigate possible effects of individual-specific characteristics of respondents on their price expectations. For example, Niu and Van Soest (2014) explore the relationship between house price expectations, local economic conditions, and individual household characteristics. Bover (2015) uses house price expectations data from the Spanish Survey of Household Finances, and finds important differences in expectations across gender and occupation. Kuchler and Zafar (2015) use data from Survey of Consumer Expectations and focus on how personal experiences affect expectations at the national level. They find that experiencing a house price fall leads respondents to be more pessimistic about future US house prices.

The above studies all point to important systematic differences in price expectations across respondents. Similar disparities in expectations are also present in our surveys. Using the information in demographic modules of ALP, we now consider the effects of sex, age, income, ethnicity and education on price expectations. Given the time-invariant nature of the demographic variables, there are two ways that this can be done. One possibility would be to augment the panel regressions in (26) with the observed individual-specific effects, and then treat $\alpha_i^{(h)}$ as random effects, distributed independently of x_{it}. Setting $\alpha_i^{(h)} = \alpha^{(h)} + \mathbf{z}_i'\boldsymbol{\gamma}^{(h)} + \psi_i^{(h)}$, where $\mathbf{z}_i$ is the vector of time-invariant observed characteristics of the i^{th} respondent, $\psi_i^{(h)}$ is the unobserved random component of $\alpha_i^{(h)}$ assumed to be distributed independently of $\mathbf{z}_i$ and x_{it}. The associated random effects panel data model can now be written as

$$\hat{\pi}^e_{i,t+h|t} = \alpha^{(h)} + \mathbf{z}_i'\boldsymbol{\gamma}^{(h)} + \beta^{(h)}x_{it} + \delta_t^{(h)} + \varepsilon_{i,t+h} + \psi_i^{(h)}. \tag{28}$$

We consider model (28) both with and without time effects $\delta_t^{(h)}$. The random effects estimator, and the random effects estimator with time dummies will be denoted by $\boldsymbol{\gamma}_{RE}^{(h)}$, $\beta_{RE}^{(h)}$ and $\boldsymbol{\gamma}_{RE-TE}^{(h)}$, $\beta_{RE-TE}^{(h)}$, respectively. For the elements of $\mathbf{z}_i = (z_{i1}, z_{i2}, ..., z_{i7})'$, we consider $z_{i1} = 1$ if the respondent identifies as female, and 0 otherwise, $z_{i2} = \ln \text{age}_i$, z_{i3} measures the education level of respondent i, $z_{i4} = \ln \text{income}_i$, and z_{i5} to z_{i7} are dummy variables that take the value of 1 if the respondent identifies her/himself as Asian, Black and Hispanic/Latino, respectively. For a detailed description of how the time-invariant variables are constructed see Appendix A.2. We allow $\varepsilon_{i,t+h} + \psi_i^{(h)}$ to be serially correlated and heteroskedastic.

An alternative approach, that does not require $\psi_i^{(h)}$ and x_{it} to be independently distributed, is to employ the two-stage approach proposed recently in Pesaran and Zhou (2016), whereby in the first stage FE (or FE-TE) estimates of $\beta^{(h)}$ are used to filter out the effects of x_{it}, and in the second stage a pure cross section regression of $\bar{\hat{u}}_i$ is run on an intercept and z_i, for $i = 1, 2, ..., N$, where

$$\bar{\hat{u}}_i = \frac{\sum_{t=1}^{T} s_{it}\left(\hat{\pi}^e_{i,t+h|t} - \hat{\beta}^{(h)}_{FE-TE}x_{it}\right)}{\sum_{t=1}^{T} s_{it}},$$

and s_{it} is an indicator variable which takes the value of 1 if respondent i is included in wave t of the survey and 0 otherwise. This estimator is referred to as the FE filtered estimator and denoted by $\hat{\boldsymbol{\gamma}}_{FEF}^{(h)}$ (or $\hat{\boldsymbol{\gamma}}_{FEF-TE}^{(h)}$). Pesaran and Zhou (2016) provide standard errors for $\hat{\boldsymbol{\gamma}}_{FEF}^{(h)}$ that allow for the sampling uncertainty of $\hat{\beta}_{FE}^{(h)}$ (or $\hat{\beta}_{FE-TE}^{(h)}$), and possible error heteroskedasticity.

The FE filtered and RE estimates of $\boldsymbol{\gamma}^{(h)}$ and their robust standard errors are summarized for equity, gold and house price expectations in Tables 8, 9 and 10, respectively. For completeness we also report the estimates of $\beta^{(h)}$, although, as noted earlier, the RE estimates are not robust to possible correlations between η_i and x_{it}. The FE estimates of $\beta^{(h)}$ in Tables 8-10 are the same as those already reported in Table 6. Inclusion of time dummies had little impact on the RE or FE estimates (the FE-TE estimates are reported in the supplement). But we find it matters a great deal, particularly to the regressions for house price expectations, if we did include a location (MSA) dummy in the regressions. As noted earlier, we have been able to identify the MSA within which a respondent resides from the demographic module of the survey and the house price information that was provided to the respondents. This additional information (often absent in other survey expectations) allows us to separate the location-specific nature of house price changes from respondent-specific characteristics.

Comparing RE and FE estimates of $\beta^{(h)}$ we note that they are generally quite close, although the RE estimates tend to be larger in absolute magnitude, and more statistically significant. Judging by the implied estimates of r, and the fact that FE estimates are robust to possible correlations between x_{it} and η_i, the FE estimates are clearly to be preferred.[11] But it is worth noting that our main conclusion that the valuation indicator plays a significant role in price expectations formation holds irrespective of whether RE or FE estimates are used. Also, RE estimates of $\beta^{(h)}$ are robust to the inclusion of location dummies.[12]

Regarding the effects of individual-specific characteristics on price expectations, we find important differences across assets. For equity prices sex, age and education are statistically significant at all three horizons and irrespective of whether RE or FE filtered estimates are considered. Ethnicity also features significantly for 3 and 12 months horizons. Females tend to have higher equity price expectations, whilst older respondents, and those with a higher level of income, tend to have lower equity price expectations. But it is interesting that the estimates and their statistical significance are hardly affected by the inclusion of location and/or time dummies (the latter results reported in the supplement). Similar results are obtained for gold price expectations where in addition to sex, age, income and ethnicity, education is also statistically significant, with higher educated respondents having lower price expectations of gold prices.

The picture is very different when we consider regressions for house price expectations (in

[11] Implied estimates of r for RE estimates are provided in the supplement.

[12] Note that the FE estimates are unaffected by respondent-specific characteristics, including their location.

Table 10). Generally speaking, the respondent-specific characteristics are not as significant as compared to the equity and gold price regressions, and the test outcomes critically depend on the estimator and whether the regressions include location dummies. Using the preferred FE filtered estimates and considering the regressions with MSA dummies, we find that only income is statistically significant (with a positive sign) in the case of regressions for one month ahead, and ethnicity for the one year expectations. The heterogeneity of house price expectations across respondents seem to be largely explained by the location dummy once we condition on the valuation indicator, and all other respondent-specific characteristics loose their statistical significance.[13]

[13] A similar result is also reported in Bover (2015) who shows that most of the observed heterogeneity in house price expectations can be explained by a location dummy at the postal code level.

Table 8: **Fixed Effect Filtered and Random Effect Estimates of Price Expectation Equations for Equity**

	One Month Ahead				Three Months Ahead				One Year Ahead			
	FEF	FEF	RE	RE	FEF	FEF	RE	RE	FEF	FEF	RE	RE
x_{it} (1)	-0.099	-0.099	-0.124	-0.133	-0.090	-0.090	-0.118*	-0.120*	-0.115***	-0.115***	-0.138***	-0.139***
	(0.127)	(0.127)	(0.116)	(0.116)	(0.076)	(0.076)	(0.0700)	(0.0703)	(0.036)	(0.036)	(0.0337)	(0.0339)
Female	0.767**	0.793***	0.654**	0.684***	0.898***	0.896***	0.825***	0.829***	0.570***	0.567***	0.553***	0.553***
	(0.301)	(0.302)	(0.261)	(0.264)	(0.171)	(0.172)	(0.155)	(0.156)	(0.084)	(0.084)	(0.0778)	(0.0783)
ln age	-2.845***	-2.718***	-2.464***	-2.345***	-2.633***	-2.568***	-2.417***	-2.358***	-1.668***	-1.628***	-1.580***	-1.545***
	(0.511)	(0.508)	(0.441)	(0.441)	(0.296)	(0.297)	(0.264)	(0.265)	(0.142)	(0.144)	(0.130)	(0.133)
Education	-0.343	-0.424	-0.305	-0.416	-0.622***	-0.690***	-0.597***	-0.680***	-0.467***	-0.502***	-0.454***	-0.494***
	(0.329)	(0.337)	(0.272)	(0.280)	(0.184)	(0.187)	(0.163)	(0.166)	(0.087)	(0.089)	(0.0800)	(0.0819)
ln income	-0.663***	-0.624**	-0.686***	-0.657***	-0.816***	-0.790***	-0.793***	-0.777***	-0.479***	-0.461***	-0.468***	-0.455***
	(0.241)	(0.247)	(0.207)	(0.213)	(0.148)	(0.151)	(0.133)	(0.135)	(0.070)	(0.071)	(0.0646)	(0.0657)
Asian	-1.577	-1.430	-1.254	-1.191	-0.210	-0.082	-0.133	-0.0571	-0.170	-0.133	-0.137	-0.115
	(1.056)	(1.061)	(0.931)	(0.948)	(0.619)	(0.629)	(0.552)	(0.568)	(0.274)	(0.280)	(0.248)	(0.258)
Black	0.956	0.787	0.998*	0.772	1.501***	1.377***	1.523***	1.369***	1.135***	1.046***	1.149***	1.051***
	(0.665)	(0.681)	(0.586)	(0.599)	(0.368)	(0.376)	(0.335)	(0.343)	(0.173)	(0.178)	(0.162)	(0.168)
Hispanic/Latino	-0.098	-0.612	0.280	-0.234	1.205***	0.928***	1.284***	1.006***	0.895***	0.759***	0.916***	0.781***
	(0.586)	(0.640)	(0.505)	(0.559)	(0.310)	(0.341)	(0.282)	(0.311)	(0.142)	(0.157)	(0.133)	(0.146)
MSA dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

The estimates reported refer to the panel regressions $\hat{\pi}^e_{i,t+h|t} = \beta^{(h)} x_{it} + \mathbf{z}_i'\boldsymbol{\gamma}^{(h)} + \alpha_i^{(h)} + \varepsilon_{i,t+h}$ using an unbalanced panel of 4,971 respondents over 11 months, March 2012 to January 2013.
$N = 35,961$, $T_{min} = 1$, $T_{p25} = 4$, $T_{p50} = 6$, $\bar{T} = 7.23$, $T_{p75} = 9$, $T_{max} = 11$
FEF - estimator of Pesaran and Zhou (2016). Standard errors are robust to heteroskedasticity and serial correlation.
RE - random effect estimates with standard errors clustered at individual level.
Standard errors are in parentheses, *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.
(1) - The FE estimates of $\beta^{(h)}$ reported in the table are the same as those summarized in Table 6.

Table 9: **Fixed Effect Filtered and Random Effect Estimates of Price Expectation Equations for Gold**

	One Month Ahead				Three Months Ahead				One Year Ahead			
	FE	FE	RE	RE	FE	FE	RE	RE	FE	FE	RE	RE
x_{it} [(1)]	0.602***	0.602***	0.581***	0.591***	0.222**	0.222**	0.208**	0.213**	-0.023	-0.023	-0.0565	-0.0524
	(0.197)	(0.197)	(0.177)	(0.178)	(0.108)	(0.108)	(0.0993)	(0.0997)	(0.049)	(0.049)	(0.0455)	(0.0457)
Female	1.174***	1.166***	1.143***	1.131***	1.062***	1.072***	1.047***	1.056***	0.622***	0.634***	0.626***	0.639***
	(0.356)	(0.358)	(0.321)	(0.323)	(0.203)	(0.202)	(0.190)	(0.190)	(0.098)	(0.098)	(0.0943)	(0.0942)
ln age	-4.132***	-4.039***	-3.925***	-3.840***	-3.140***	-3.117***	-3.072***	-3.052***	-1.725***	-1.722***	-1.707***	-1.708***
	(0.591)	(0.595)	(0.539)	(0.542)	(0.332)	(0.334)	(0.313)	(0.314)	(0.158)	(0.159)	(0.152)	(0.152)
Education	-1.312***	-1.277***	-1.294***	-1.266***	-1.280***	-1.235***	-1.269***	-1.226***	-0.798***	-0.780***	-0.783***	-0.768***
	(0.370)	(0.387)	(0.319)	(0.330)	(0.220)	(0.223)	(0.198)	(0.200)	(0.103)	(0.104)	(0.0965)	(0.0979)
ln income	-1.362***	-1.302***	-1.381***	-1.325***	-1.119***	-1.070***	-1.140***	-1.092***	-0.656***	-0.624***	-0.663***	-0.632***
	(0.302)	(0.309)	(0.265)	(0.271)	(0.172)	(0.175)	(0.158)	(0.161)	(0.080)	(0.080)	(0.0750)	(0.0751)
Asian	0.981	1.046	0.758	0.833	0.556	0.651	0.492	0.590	0.299	0.389	0.274	0.365
	(1.407)	(1.421)	(1.212)	(1.232)	(0.687)	(0.706)	(0.651)	(0.666)	(0.325)	(0.338)	(0.318)	(0.327)
Black	1.977***	1.639**	2.071***	1.742**	1.886***	1.756***	1.898***	1.768***	1.309***	1.254***	1.297***	1.240***
	(0.751)	(0.769)	(0.695)	(0.711)	(0.409)	(0.414)	(0.388)	(0.394)	(0.198)	(0.201)	(0.189)	(0.193)
Hispanic/Latino	1.280**	0.645	1.452**	0.856	1.682***	1.339***	1.674***	1.355***	0.967***	0.802***	0.954***	0.800***
	(0.640)	(0.717)	(0.573)	(0.645)	(0.353)	(0.395)	(0.327)	(0.367)	(0.161)	(0.179)	(0.153)	(0.170)
MSA dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

The estimates reported refer to the panel regressions $\hat{\pi}^{e}_{i,t+h|t} = \beta^{(h)} x_{it} + \mathbf{z}_i' \boldsymbol{\gamma}^{(h)} + \alpha_i^{(h)} + \varepsilon_{i,t+h}$ using an unbalanced panel of 4,971 respondents over 11 months, March 2012 to January 2013.
$N = 35,961$, $T_{min} = 1$, $T_{p25} = 4$, $T_{p50} = 6$, $\bar{T} = 7.23$, $T_{p75} = 9$, $T_{max} = 11$
FEF - estimator of Pesaran and Zhou (2016). Standard errors are robust to heteroskedasticity and serial correlation.
RE - random effect estimates with standard errors clustered at individual level.
Standard errors are in parentheses, *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.
[(1)] - The FE estimates of $\beta^{(h)}$ reported in the table are the same as those summarized in Table 6.

Table 10: **Fixed Effect Filtered and Random Effect Estimates of Price Expectation Equations for Housing**

	One Month Ahead				Three Months Ahead				One Year Ahead			
	FE	FE	RE	RE	FE	FE	RE	RE	FE	FE	RE	RE
x_{it} [(1)]	-0.292***	-0.292***	-0.363***	-0.382***	-0.106***	-0.106***	-0.144***	-0.149***	-0.048***	-0.048***	-0.0638***	-0.0634***
	(0.064)	(0.064)	(0.0604)	(0.0547)	(0.027)	(0.027)	(0.0254)	(0.0242)	(0.010)	(0.010)	(0.00945)	(0.00940)
Female	0.139	-0.000	0.123	-0.0138	0.032	-0.016	0.0236	-0.0240	0.037	0.023	0.0356	0.0216
	(0.185)	(0.107)	(0.180)	(0.092)	(0.069)	(0.047)	(0.0666)	(0.0426)	(0.026)	(0.023)	(0.0250)	(0.0216)
ln age	2.041***	0.046	2.007***	-0.0146	0.673***	0.002	0.663***	-0.00791	0.189***	0.024	0.187***	0.0227
	(0.287)	(0.162)	(0.280)	(0.134)	(0.107)	(0.073)	(0.103)	(0.0632)	(0.040)	(0.037)	(0.0380)	(0.0334)
Education	0.389**	0.086	0.384**	0.114	0.165**	0.044	0.166**	0.061	0.042*	0.006	0.0428*	0.00841
	(0.187)	(0.121)	(0.181)	(0.102)	(0.068)	(0.049)	(0.0652)	(0.0439)	(0.025)	(0.022)	(0.0237)	(0.0211)
ln income	0.549***	0.248***	0.546***	0.247***	0.145***	0.047	0.145***	0.0497	0.011	-0.012	0.0137	-0.00911
	(0.135)	(0.087)	(0.132)	(0.0733)	(0.052)	(0.040)	(0.0502)	(0.0353)	(0.021)	(0.019)	(0.0196)	(0.0179)
Asian	-1.300*	-0.097	-1.332*	-0.233	-0.448*	-0.041	-0.443*	-0.0547	-0.016	0.067	-0.00995	0.0673
	(0.738)	(0.443)	(0.716)	(0.393)	(0.246)	(0.159)	(0.242)	(0.154)	(0.089)	(0.076)	(0.0879)	(0.0753)
Black	-1.490***	-0.289	-1.429***	-0.210	-0.433***	-0.035	-0.414***	-0.0211	-0.040	0.054	-0.0362	0.0548
	(0.354)	(0.222)	(0.345)	(0.192)	(0.133)	(0.100)	(0.128)	(0.0889)	(0.053)	(0.049)	(0.0500)	(0.0456)
Hispanic/Latino	-2.033***	-0.165	-2.074***	-0.196	-0.500***	0.119	-0.521***	0.100	-0.004	0.135***	-0.0114	0.130***
	(0.295)	(0.185)	(0.290)	(0.159)	(0.114)	(0.088)	(0.111)	(0.0783)	(0.043)	(0.042)	(0.0412)	(0.0390)
MSA dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

The estimates reported refer to the panel regressions $\hat{\pi}^e_{i,t+h|t} = \beta^{(h)} x_{it} + \mathbf{z}_i' \boldsymbol{\gamma}^{(h)} + \alpha_i^{(h)} + \varepsilon_{i,t+h}$ using an unbalanced panel of 4,971 respondents over 11 months, March 2012 to January 2013.

$N = 35,961$, $T_{min} = 1$, $T_{p25} = 4$, $T_{p50} = 6$, $\bar{T} = 7.23$, $T_{p75} = 9$, $T_{max} = 11$

FEF - estimator of Pesaran and Zhou (2016). Standard errors are robust to heteroskedasticity and serial correlation.

RE - random effect estimates with standard errors clustered at individual level.

Standard errors are in parentheses, *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.

[(1)] - The FE estimates of $\beta^{(h)}$ reported in the table are the same as those summarized in Table 6.

2

5 Bubbles and crashes: leading indicators

The equilibrium relation between expected price changes and the valuation indicator in (20) can also be used to construct time series indicators of bubbles and crashes at the level of individual respondents, that can then be aggregated to regional or national levels. Such indicators are likely to provide valuable information about the possibility of bubbles or crashes building up. In what follows we suggest such indicators. But since the survey results are available only over a very short time period, a time series evaluation of the usefulness of such indicators is not possible. Instead we consider a related question of whether spatially disaggregated bubble and crash indicators can help explain the cross-section variations of realized house price changes across MSAs. Specifically, we consider 48 MSAs that have at least 20 respondents on average during the 11 survey months.

We begin with respondent-specific indicators and consider the conjunctions of individual i^{th} responses to the double-question surveys that contradict the theoretical relations between $\hat{\pi}^e_{i,t+h|t}$ and x_{it}, namely when respondent's valuation belief and price change expectations do not match. Accordingly, we define the bubble indicator for respondent i at time t for h periods ahead by

$$B_{i,t+h|t} = I[(x_{it} > 0 \; and \; \hat{\pi}^e_{i,t+h|t} \geq 0) \; or \; (x_{it} = 0 \; and \; \hat{\pi}^e_{i,t+h|t} > 0)], \tag{29}$$

and the crash indicator by

$$C_{i,t+h|t} = I[(x_{it} < 0 \; and \; \hat{\pi}^e_{i,t+h|t} \leq 0) \; or \; (x_{it} = 0 \; and \; \hat{\pi}^e_{i,,t+h|t} < 0)]. \tag{30}$$

By implication we consider a respondent as being neutral if their price change expectations, $\hat{\pi}^e_{i,t+h|t}$, and valuation belief indicator, x_{it}, is in accordance with the theoretical relationship given by (20).

The average bubble indicator for MSA s at time t for h periods ahead is then defined by

$$B_{s,t+h|t} = \frac{\sum_{i_t \in \Theta_{st}} B_{i,t+h|t}}{\#\Theta_{st}}, \tag{31}$$

where Θ_{st} denotes the set of respondents in MSA s at time t. Similarly, the average crash indicator for MSA s at time t for h periods ahead is defined by

$$C_{s,t+h|t} = \frac{\sum_{i_t \in \Theta_{st}} C_{i,t+h|t}}{\#\Theta_{st}}. \tag{32}$$

For each MSA s, we also define average bubble and crash indicators of neighboring areas as follows. Let $\mathbf{W} = \{w_{ss'}\}_{s,s'=1,2,\ldots,N}$ denote an $N \times N$ matrix with $w_{ss'} = 1$ if MSAs s and s' lie in neighboring areas, and $w_{ss'} = 0$, otherwise. Specifically, for each MSA s we consider the Haversine distance between its geographic center and that of other MSAs, denoted by $\rho(s, s')$ and measured in miles. If there is at least one MSA s' such that $\rho(s, s') \leq 100$, then we set $w_{ss'} = 1$ for all $s' \neq s$ such that $\rho(s, s') \leq 100$, and 0, otherwise. If there are no MSAs within the 100 mile radius, then we repeat the exercise with a 200 miles radius and set $w_{ss'} = 1$ for all MSAs s' that satisfy $\rho(s, s') \leq 200$. Finally, if there are no MSAs within 200 miles of MSA s, we treat all other MSAs as neighboring areas, i.e. $w_{ss'} = 1$ for all $s' \neq s$ (this was the case for 13 out of 48 MSAs).[14] The average neighboring area bubble and crash indicators for MSA s in month t are defined by

$$B^*_{s,t+h|t} = \frac{\sum_{s'=1}^{N} w_{ss'} B_{s,t+h|t}}{\sum_{s'=1}^{N} w_{ss'}}, \tag{33}$$

and

$$C^*_{s,t+h|t} = \frac{\sum_{s'=1}^{N} w_{ss'} C_{s,t+h|t}}{\sum_{s'=1}^{N} w_{ss'}}. \tag{34}$$

To visualize the bubble and crash indicators, we can summarize the responses in 3×3 contingency tables of beliefs about current prices (rows) against the expected future price changes (columns). For a given MSA s, month t and expectation horizon h, let $N_{ij,s,t+h|t}$ denote the number of responses in category (i, j), with $i = u$ (under-valued), f (fair), o (over-valued), $j = r$ (rise), s (same), and f (fall).

The bubble and crash indicators are then computed as

$$B_{s,t+h|t} = \frac{N_{fr,s,t+h|t} + N_{or,s,t+h|t} + N_{os,s,t+h|t}}{\sum_{i,j} N_{ij,s,t+h|t}},$$

[14] We assume $w_{ss'} = 0$ for all $s = s'$. A detailed description of how the spatial matrix is calculated can be found in Appendix A.3.

Table 11: **Valuation-expectations response categories**

Current valuation	**Expected future price change** (a) Up	(b) Same	(c) Down
(a) Under-valued	$N_{ur,s,t+h\|t}$	$N_{us,s,t+h\|t}$	$N_{uf,s,t+h\|t}$
(b) Fairly-valued	$N_{fr,s,t+h\|t}$	$N_{fs,s,t+h\|t}$	$N_{ff,s,t+h\|t}$
(c) Over-valued	$N_{or,s,t+h\|t}$	$N_{os,s,t+h\|t}$	$N_{of,s,t+h\|t}$

and

$$C_{s,t+h|t} = \frac{N_{us,s,t+h|t} + N_{uf,s,t+h|t} + N_{ff,s,t+h|t}}{\sum_{i,j} N_{ij,s,t+h|t}}.$$

6 Explanation of realized house price changes using bubble and crash indicators

In what follows we empirically investigate the value added of the above average bubble and crash indicators, as well as spillover effects captured by their neighboring area counterparts, in explaining realized house price changes across the 48 MSAs over the 11 survey waves. As a bench mark model we consider the following standard dynamic panel regression model for expectation horizons $h = 1, 3, 12$ months.

$$M_1 : \pi_{s,t+1} = \alpha_s^{(h)} + \lambda_0^{(h)} \pi_{st} + \lambda_1^{(h)} \hat{\pi}^e_{s,t+h|t} + u_{s,t+1,h}, \text{ for } h = 1, 3, 12, \tag{35}$$

where $\pi_{s,t+1} = 300 \left[\ln(P_{s,t+1}) - \ln(P_{st})\right]$ is the one month ahead realized house price change in MSA s (expressed in per cent per quarter), and $\hat{\pi}^e_{s,t+h|t}$ is the expected house price change formed in month t for h months ahead, and averaged across the respondents in MSA s.[15] Specifically

$$\hat{\pi}^e_{s,t+h|t} = \frac{\sum_{i_t \in \Theta_{st}} \hat{\pi}^e_{i,t+h|t}}{\#\Theta_{st}}.$$

[15] Note that $\hat{\pi}^e_{s,t+h|t} = S_h \left[\ln\left(P^e_{s,t+h|t}\right) - \ln\left(P_{st}\right)\right]$, for $h = 1, 3, 12$, with $S_1 = 300$, $S_3 = 100$ and $S_{12} = 25$.

Given the importance of location in the formation of house price expectations discussed above, we also allow for MSA-specific fixed effects, $\alpha_s^{(h)}$, in the benchmark model. We then augment the benchmark model (35), with the MSA-specific bubble and crash indicators. We consider the following specification

$$\begin{aligned} M_2 \quad : \quad \pi_{s,t+1} &= \alpha_s^{(h)} + \lambda_0^{(h)}\pi_{st} + \lambda_1^{(h)}\hat{\pi}^e_{s,t+h|t} + \beta_1^{(h)} B_{s,t+h|t} + \beta_2^{(h)} C_{s,t+h|t} \\ &+ \gamma_1^{(h)} B^*_{s,t+h|t} + \gamma_2^{(h)} C^*_{s,t+h|t} + u_{s,t+1,h}. \end{aligned} \tag{36}$$

To isolate the importance of the bubble and crash indicators from the price expectations we also estimate (36), without the expectations variable, $\hat{\pi}^e_{s,t+h|t}$, which we denote as model M_3.

All three specifications are estimated using a balanced panel of observations over $N = 48$ MSAs, and $T = 9$ months, namely for $s = 1, 2, \ldots, 48$, and $t =$ May 2012 - January 2013. First-differencing is applied to eliminate the MSA-specific effects. Note that standard FE estimation of dynamic panel regressions will not be appropriate since T is small relative to N, and FE estimates can lead to significant bias due to the presence of the lagged dependent variable in the panel regressions. After first-differencing we estimate the parameters by the two-step Generalized Method of Moments (GMM) method due to Arellano and Bond (1991), using the following moment conditions:[16]

$$E\left(\Delta u_{s,t+1,h}\mathbf{z}_{s,j}\right) = \mathbf{0}, \text{ for } j = t-2, t-1; t = 3(\text{May 2012}), 5, ..., 11(\text{January 2013}); \tag{37}$$

where we set $\mathbf{z}_{s,j} = \left(\pi_{s,j}, \hat{\pi}^e_{s,j+h|j}\right)'$, for the baseline model M_1,

$$\mathbf{z}_{s,j} = \left(\pi_{s,j}, \hat{\pi}^e_{s,j+h|j}, B_{s,j+h|j}, C_{s,j+h|j}, B^*_{s,j+h|j}, C^*_{s,j+h|j}\right)', \text{ for model } M_2,$$

and

$$\mathbf{z}_{s,j} = \left(\pi_{s,j}, B_{s,j+h|j}, C_{s,j+h|j}, B^*_{s,j+h|j}, C^*_{s,j+h|j}\right)', \text{ for model } M_3.$$

The estimation results are summarized in Table 12. Note that we are primarily interested in the explanatory power of house price inflation expectations, $\hat{\pi}^e_{s,t+h|t}$, and the crash and

[16] Note that we do not use all available moment conditions suggested by Arellano and Bond (1991), to avoid the weak instrument problem.

bubble indicators $B_{s,t+h|t}$, $C_{s,t+h|t}$, $B^*_{s,t+h|t}$, and $C^*_{s,t+h|t}$. The lagged value of realized house price changes, π_{st}, is included in the analysis to take account of the high degree of known persistence in realized price changes. Consider first the estimates for the baseline model, M_1. As expected, $\lambda_0^{(h)}$ which measure the degree of persistence in the rate of house price changes, is estimated to be quite high and lies in the range $0.70 - 0.80$, and is statistically significant at all horizons. The coefficient of house price expectations formed at t, $\lambda_1^{(h)}$, is also statistically significant but its magnitude is disappointingly low, and in fact becomes negative for $h = 12$. In contrast, the bubble and crash indicators, included in model M_2, are statistically significant and have the correct signs for all horizons, $h = 1, 3$, and 12. For $h = 1$, the panel regressions predict that MSAs with a higher bubble indicator tend to experience a higher degree of house price changes, and MSAs with a higher crash indicator tend to experience a lower degree of house price changes. It is also most interesting that similar effects are observed from spillover bubble and crash indicators, in the sense that MSAs that are surrounded by neighboring MSAs with a high (low) value of the bubble (crash) indicator also tend to show a higher (lower) degree of house price changes. The effects of changes in bubble and crash indicators on future house price changes get accentuated due to the fact that in general the bubble and crash indicators move in opposite directions. Finally, these results continue to hold even if the price expectations variable is dropped from the analysis. See the estimates under columns M_2 and M_3 in Table 12.

Overall, the bubble and crash indicators and the associated neighboring indicators seem to play an important role in future movements of realized house price changes across MSAs. For example, the estimates of model M_2 for the one month expectation horizon imply that an increase in the bubble indicator from 0.2 to 0.5 leads to a 0.87 percentage point increase in the quarterly growth rate of house prices. However, due to the short nature of the time period of the surveys, a time series analysis of the out-of-sample predictive value of the crash and bubble indicators is not possible. Prediction of MSA-specific house price changes is also complicated due to the unobserved fixed effects, $\alpha_s^{(h)}$, which cannot be estimated consistently when T is short.

Table 12: **Dynamic panel regressions of realized house prices in terms of bubble and crash indicators (Across 48 MSAs and months March 2012 to January 2013)**
Dependent variable: $\pi_{s,t+1}$ **(in percent per quarter)**

	One Month ($h = 1$)			Three Months ($h = 3$)			One Year ($h = 12$)		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
π_{st}	0.712***	0.798***	0.801***	0.704***	0.783***	0.778***	0.721***	0.747***	0.744***
	(0.00872)	(0.00730)	(0.00554)	(0.00772)	(0.00605)	(0.00628)	(0.00528)	(0.00752)	(0.00496)
$\hat{\pi}^e_{s,t+h\|t}$	0.0159***	-0.0505***		0.0513***	-0.111***		-0.0924***	-0.292***	
	(0.00231)	(0.00783)		(0.00697)	(0.0187)		(0.0217)	(0.0390)	
$B_{s,t+h\|t}$		2.905***	2.055***		3.883***	3.316***		2.307***	1.613***
		(0.378)	(0.635)		(0.622)	(0.568)		(0.347)	(0.232)
$C_{s,t+h\|t}$		-5.886***	-4.653***		-3.261***	-1.518**		-3.379***	-2.657***
		(0.500)	(0.446)		(0.685)	(0.705)		(0.983)	(0.530)
$B^*_{s,t+h\|t}$		7.016***	7.861***		10.12***	11.01***		3.389***	3.839***
		(0.778)	(0.826)		(0.487)	(0.252)		(0.552)	(0.384)
$C^*_{s,t+h\|t}$		-5.310***	-5.409***		-0.722	-0.576		-3.755***	-3.437***
		(0.681)	(0.503)		(0.486)	(0.602)		(1.028)	(0.728)

The panel regression is estimated using a two-step GMM estimator (Arellano and Bond (1991)) using the moment conditions specified in Section 6 with heteroskedasticity-robust standard errors.
Observations from the first two survey waves March to April 2012 are used to initialize moment conditions.
The estimates are based on a balanced panel with $N = 48$ and $T = 9$.
Standard errors are in parentheses, *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.

7 Concluding remarks

In this paper we have introduced a new type of survey which combines standard surveys of price expectations with questions regarding the respondents' subjective belief about asset values. Using a theoretical asset pricing model with heterogenous agents we show that there exists a negative relationship between the agents expectations of price changes and their asset valuation, a relationship that holds under different horizons. Double-question surveys provide evidence in support of such relationships, particularly for house prices for which survey respondents are more likely to have a first-hand knowledge as compared to other assets such as equities or gold prices which might not be of concern to many respondents in the survey. We also investigate the effects of demographic factors, such as sex, age, education, ethnicity, and income on price expectations, and find important differences in price expectations. But, interestingly enough, for house price expectations demographic factors stop being statistically significant once we condition on the respondent's location and his/her valuation indicator. Finally, we show how the results of the double-question surveys can be used to construct leading bubble and crash indicators for use in forecasting and policy analyses. The potential value of such indicators is illustrated in a dynamic panel regression of realized house price changes across a number of key MSAs in the US.

We consider the double-question surveys carried out so far, and the analysis of the survey results that we have provided, as a prototype study which needs to be pursued further by government and international agencies, particularly central banks. It is only by further critical analysis and the conduct of similar surveys in the US and elsewhere that the true worth of results from double-question surveys as leading indicators of bubbles and crashes can be ascertained.

References

Abreu, D. and M. K. Brunnermeier (2003). Bubbles and crashes. *Econometrica 71*(1), 173–204.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Blanchard, O. J. (1979). Speculative bubbles, crashes and rational expectations. *Economics letters 3*(4), 387–389.

Bover, O. (2015). Measuring expectations from household surveys: new results on subjective probabilities of future house prices. *SERIEs 6*(4), 361–405.

Branch, W. A. (2004). The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *The Economic Journal 114*(497), 592–621.

Case, K. E. and R. J. Shiller (1988). The behavior of home buyers in boom and post-boom markets. Working Paper 2748, National Bureau of Economic Research.

Case, K. E. and R. J. Shiller (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity 2003*(2), 299–362.

Case, K. E., R. J. Shiller, and A. Thompson (2012). What have they been thinking? home buyer behavior in hot and cold markets. Working Paper 18400, National Bureau of Economic Research.

Chen, J., H. Hong, and J. C. Stein (2002). Breadth of ownership and stock returns. *Journal of financial Economics 66*(2), 171–205.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance 53*(6), 1839–1885.

Dominitz, J. and C. F. Manski (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics 26*(3), 352–370.

Flood, R. P. and P. M. Garber (1980). Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy 88*(4), 745–770.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), 393–408.

Gürkaynak, R. S. (2008). Econometric tests of asset price bubbles: taking stock. *Journal of Economic Surveys 22*(1), 166–186.

Harrison, J. M. and D. M. Kreps (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics 92*(2), 323–336.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance 56*(4), 1533–1597.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Hommes, C. H. (2006). Heterogeneous agent models in economics and finance. Volume 2 of *Handbook of Computational Economics*, Chapter 23, pp. 1109 – 1186. Elsevier, Holland.

Ito, T. (1990). Foreign exchange rate expectations: Micro survey data. *The American Economic Review 80*(3), 434–449.

King, G., C. J. Murray, J. A. Salomon, and A. Tandon (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American political science review 98*(01), 191–207.

Kuchler, T. and B. Zafar (2015). Personal experiences and expectations about aggregate outcomes. *Federal Reserve Bank of New York Staff Reports* (748).

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance 32*(4), 1151–1168.

Niu, G. and A. Van Soest (2014). House price expectations. *IZA Discussion Paper* (8536).

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance 53*(6), 1887–1934.

Pesaran, M. (1987). *The Limits to Rational Expectations.* Oxford: Basil Blackwell.

Pesaran, M. H. and M. Weale (2006). Survey expectations. Volume 1 of *Handbook of economic forecasting*, Chapter 14, pp. 715–776. Elsevier, Holland.

Pesaran, M. H. and Q. Zhou (2016). Estimation of time-invariant effects in static panel data models. *Econometric Reviews (forthcoming)*.

Phelps, E. (1983). The trouble with rational expectations and the problem of inflation stabilization. Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined, Chapter 2, pp. 31–46. Cambridge University Press, Cambridge.

Phillips, P. C., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1078.

Phillips, P. C., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Scheinkman, J. A. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1220.

Shiller, R. (2000). *Irrational exuberance.* Princeton, NJ: Princeton University Press, Princeton.

Shleifer, A. (2000). *Clarendon Lectures: Inefficient Markets.* Oxford University Press, New York. Japanese translation, Toyo Keisai, Tokyo, 2001. Chinese translation, 2003.

Thaler, R. H. (1991). *Quasi Rational Economics* (1 ed.). Russell Sage Foundation, New York.

Townsend, R. M. (1983). Forecasting the forecasts of others. *Journal of Political Economy 91*(4), 546–588.

A Appendix

A.1 Survey questions

We are interested in learning your views about prices of houses, stocks and shares, and gold, and appreciate your responses to the following questions.

H1 rate current housing prices
We now have some questions about housing prices. The median price of a single family home in the [*fill for city nearest to R zip code*] cosmopolitan area is currently around [*converted fill for median housing price in R zip code area*] (Half of all single family homes in the area cost less than the median, and the other half cost more than the median.). Do you believe that current housing prices are:
1 just right (in the sense that housing prices are in line with what you personally regard to be fair),
2 too high,
3 too low as compared to the fair value?

H2_intro
Bearing in mind your response to the previous question, suppose now that someone were to purchase a single family home in [*fill for city nearest to R zip code*] area for the price of [. . .] What do you expect the house to be worth (Please enter a numeric answer only, with no commas or punctuation)
H2_1month 1 month from now,
H2_3month 3 months from now,
H2_1year 1 year from now.

Respondents who reside further than 500 miles away from a major metropolitan area were provided with **H1_alternate** and **H2_intro_alternate** instead of **H1 and H2_intro.**

H1_alternate rate current housing prices
We now have some questions about housing prices. The median price of a single family home in the USA is currently around $163,500 (Half of all single family homes in the area cost

less than the median, and the other half cost more than the median.). Do you believe that current housing prices are:
1 just right (in the sense that housing prices are in line with what you personally regard to be fair),
2 too high,
3 too low as compared to the fair value?

H2_intro_alternate
Bearing in mind your response to the previous question, suppose now that someone were to purchase a single family home in the USA for the price of $163,500. What do you expect the house to be worth (Please enter a numeric answer only, with no commas or punctuation)
H2_1month 1 month from now,
H2_3month 3 months from now,
H2_1year 1 year from now.

H3_intro
Will you please elaborate by providing responses to the following: What do you think is the per cent chance that one year from now the house will be worth
H3_percent1 amount minus or plus 5 per cent. Between [*calculated low house value*] and [*calculated high house value*] dollars?
H3_percent2 amount less 5 per cent. Less than [*calculated low house value*] dollars?
H3_percent3 amount more than 5 per cent. More than [*calculated high house value*] dollars?
Your responses should add up to 100 per cent.

E1 rate stock price level
We have some questions about the price of publicly traded stocks. Do you believe the US stock market (as measured by S&P 500 index) to be currently:
1 Overvalued
2 Fairly valued (in the sense that the general level of stock prices is in line with what you personally regard to be fair)
3 Undervalued

E1_note explain stock index

Note: The S&P 500 is an index of 500 common stocks actively traded in the United States. It provides one measure of the general level of stock prices.

E2_intro estimate 1000 investment

Bearing in mind your response to the previous question, suppose now that today someone were to invest 1000 dollars in a mutual fund that tracks the movement of S&P 500 very closely. That is, this "index fund" invests in shares of the companies that comprise the S&P 500 Index. What do you expect the $1000 investment in the fund to be worth

E2_1month in one month from now,

E2_3month in three months from now,

E2_1year in one year from now.

E3_intro intro to per cent change

Will you please elaborate by providing responses to the following: What do you think is the per cent chance that a year from today the investment will be worth

E3_percent1 minus 5 to plus 5 per cent. Between [*calculated low stock value*] and [*calculated high stock value*] dollars?

E3_percent2 minus 5 per cent. Less than [*calculated low stock value*] dollars?

E3_percent3 plus 5 per cent. More than [*calculated high stock value*] dollars?

Your responses should add up to 100 **per cent**.

G1 rate current gold prices

We now have some questions about the price of gold bullion traded internationally. Given the current price of gold, do you believe gold prices to be:

1 Overvalued

2 Fairly valued (in the sense that the general level of stock prices is in line with what you personally regard to be fair)

3 Undervalued

G2_intro intro to G2

Bearing in mind your response to the previous question, suppose now that today someone were to invest 1000 dollars in gold bullion. What do you expect the $1000 investment in gold to be worth
G2_1month 1 month from now,
G2_3month 3 months from now,
G2_1year 1 year from now.

G3_intro intro to G3
Will you please elaborate by providing responses to the following: What do you think is the per cent change that a year from today the investment in gold will be worth
G3_percent1 minus 10 to plus 10 per cent. Between [*calculated low gold value*] and [*calculated high gold value*] dollars?
G3_percent2 minus 10 per cent. Less than [*calculated low gold value*] dollars?
G3_percent3 plus 10 per cent. More than [*calculated high gold value*] dollars?
Your responses should add up to 100 per cent.

A.2 Truncation filters

Denote the price of asset a, with $a = eq, gd, hs$ (equity, gold, house), provided to respondent i at time t by $P_{it}^{(a)}$. Note that $P_{it}^{(eq)} = 1000$ and $P_{\ell t}^{(gd)} = 1000$, for all t. The price of asset a expected by the i^{th} respondent in month t for h months ahead is denoted by $P_{i,t+h|t}^{e,(a)}$. Respondent i's subjective valuation of asset a in period t is denoted by $x_{it}^{(a)}$, with $x_{it}^{(a)} = 1$ if the respondent believes that the asset is over-valued, $x_{it}^{(a)} = -1$ if the respondent believes that the asset is under-valued, and $x_{it}^{(a)} = 0$, otherwise.

$\mathbf{z}_i$ is a 7×1 vector of time-invariant characteristics of the i^{th} respondent. Let $\mathcal{T}_i$ be the set of time periods (months) in which respondent i takes part in the survey. The elements of $\mathbf{z}_i$ are

- $z_{i1} = 1$ if female, 0 otherwise.
- $z_{i2} = \frac{1}{\#\mathcal{T}_i} \sum_{t \in \mathcal{T}_i} \log age_{it}$, average log age of respondent i.

- $z_{i3} = \frac{1}{\#\mathcal{T}_i}\sum_{t\in\mathcal{T}_i} edu_{it}$ respondent's education averaged over the time period the respondent participated in the survey, where $edu_{it} = 0$ if the respondent has no high school diploma, $edu_{it} = 1$ if the respondent is a high school graduate with a diploma, some college but no degree, an associate degree in college occupational/vocational or academic program, and $edu_{it} = 2$ if the respondent has a Bachelor's degree or higher.[17]

- $z_{i4} = \frac{1}{\#\mathcal{T}_i}\sum_{t\in\mathcal{T}_i} \log income_{it}$, average log income of respondent i.

- $z_{i5} = 1$ if Asian, 0 otherwise.

- $z_{i6} = 1$ if Black, 0 otherwise.

- $z_{i7} = 1$ if Hispanic/Latino, 0 otherwise.

We came across a few cases where responses to gender and ethnicity questions did not remain invariant over the different survey waves. In such cases we used the following rule. Let d_{it} be the binary variable that denotes the gender or ethnicity (Asian, Black, Hispanic/Latino) of respondent i in month t, and let $\mathcal{T}_i$ denote the set of months during which respondent i participated in the surveys. Let $\bar{d}_i = \frac{1}{\#\mathcal{T}_i}\sum_{t\in\mathcal{T}_i} d_{it}$. If d_{it} varies over time, we consider the following cases.

- If $\bar{d}_i \geq 2/3$, we set $d_{it} = 1$ for all $t \in \mathcal{T}_i$.

- If $\bar{d}_i \leq 1/3$, we set $d_{it} = 0$ for all $t \in \mathcal{T}_i$.

- If $1/3 < \bar{d}_i < 2/3$, we remove respondent i from the data.

A.2.1 Truncation filter criteria

For respondent i in period t, $x_{it}^{(a)}$, $P_{i,t+h|t}^{e,(a)}$ for $a = eq, gd, hs$, and $h = 1, 3, 12$, are removed from the data set if any of the following criteria apply:

(a) **Missing responses**

- $x_{it}^{(a)}$ or $P_{i,t+h|t}^{e,(a)}$ is missing for any $a = eq, gd, hs$ or any $h = 1, 3, 12$.

[17] $z_{5,i}$ $z_{6,i}$ and $z_{7,i}$ are constructed *after* all steps of the truncation filter described in Section A.2.1 have been applied.

- $z_{1,i}$, $z_{2,i}$, $z_{3,i}$, $z_{4,i}$, age_{it}, $income_{it}$ or edu_{it} are missing.

(b) **Equity prices**

- $P^{e,(eq)}_{i,t+h|t} > 4000$ for any $h = 1, 3, 12$,
- $P^{e,(eq)}_{i,t+h|t} < 100$ for all h,
- $P^{e,(eq)}_{i,t+h|t} > 2000$ for all h,
- $P^{e,(eq)}_{i,t+h|t} = 0$ for any $h = 1, 3, 12$.

Examples of responses ($P^{e,(eq)}_{i,t+1|t}$, $P^{e,(eq)}_{i,t+3|t}$, $P^{e,(eq)}_{i,t+12|t}$) that would be truncated are: $(4020, 1030, 1020)$, $(90, 80, 99)$, $(2020, 2010, 3000)$. Examples of responses that would not be truncated are $(90, 1020, 1010)$, $(2030, 2020, 1050)$.

(b) **Gold prices**

- $P^{e,(gd)}_{i,t+h|t} > 4000$ for any $h = 1, 3, 12$
- $P^{e,(gd)}_{i,t+h|t} < 100$ for all h,
- $P^{e,(gd)}_{i,t+h|t} > 2000$ for all h,
- $P^{e,(gd)}_{i,t+h|t} = 0$ for any $h = 1, 3, 12$.

(a) **House prices**

- $P^{e,(hs)}_{i,t+h|t} < 0.5P^{(hs)}_{it}$ for any $h = 1, 3, 12$,
- $P^{e,(hs)}_{i,t+h|t} > 2P^{(hs)}_{it}$ for any $h = 1, 3, 12$,
- $P^{e,(hs)}_{i,t+h|t} = 0$ for any $h = 1, 3, 12$.

A.3 Spatial weight matrix

Consider MSAs $s = 1, 2, \ldots, S$. Let $G^{(d)}$ denote the $S \times S$ geodesic based spatial matrix calculated using the Haversine distance between MSAs. Specifically, we say that MSA s and s' are d-neighbors if the Haversine distance between their geographic centers is less than or equal to d miles. Then $G^{(d)}(s, s') = 1$ if s and s' are d-neighbors, and $G^{(d)}(s, s') = 0$ otherwise. Also, $G^{(d)}(s, s) = 0$ for all $s = 1, 2, \ldots, S$.
Denote the s^{th} row of a matrix $\mathbf{A}$ by $[\mathbf{A}]_s$ and let $a_{ss'}$ denote the (s, s') element of $\mathbf{A}$, and let $\mathbf{0}_S$ be a $1 \times S$ vector of zeros, and define $\mathbf{W} = (w_{ss'})$ as follows**.** For $s = 1, 2, \ldots, S$,

- $[\mathbf{W}]_s = [\mathbf{G}^{(100)}]_s$ if $[\mathbf{G}^{(100)}]_s \neq \mathbf{0}_S$.
- If $[\mathbf{G}^{(100)}]_s = \mathbf{0}_S$ and $[\mathbf{G}^{(200)}]_s \neq \mathbf{0}_S$, $[\mathbf{W}]_s = [\mathbf{G}^{(200)}]_s$.
- If $[\mathbf{G}^{(200)}]_s = \mathbf{0}_S$, $w_{ss'} = 1$ for $s' = 1, 2, \ldots, S,\ s' \neq s$ and $w_{ss} = 0$.