Caught in the Housing Crash: Model Failure or Management Failure?

Gunter Löffler*

University of Ulm

December 2008

Abstract

I apply standard time series models to US housing prices. Forecasts made in 2005 or earlier would have produced stress scenarios that are worse than the subsequent actual change in housing prices. The probability of these scenarios is in the range that banks claim to consider in their risk management. Hence, the fact that the crash caught many financial institutions by surprise should not be attributed to deficiencies in the traditional risk modeling approach. It seems instead that risk managers failed to apply their toolboxes, or that bank managers overruled their risk managers' assessments.

JEL classification: C22, C53, G32

Key words: housing crash, risk management, forecasting, stress scenario, ARIMA

* Gunter Löffler, Professor of Finance, Department of Finance, University of Ulm, Helmholtzstrasse 18, 89069 Ulm, Germany. Phone: ++49-731-5023597. Fax: ++49-731-5023950. E-mail: gunter.loeffler uni-ulm.de

1 Introduction

Many factors have contributed to the recent financial crisis. The surge and subsequent decline in US real estate prices is one of the most important ones. By itself, a housing crash is not enough to spark a banking crisis. If the estimated probability of a crash is large enough, financial institutions should either shy away from the market or set aside enough capital to weather a storm.[1] Surely, many financial institutions failed to do this.

But who takes the blame? Some observers endorse the view that risk models did not produce the right forecasts. Brunnermeier (2008, p. 8), for example, states, "[…] the statistical models of many professional investors and credit-rating agencies provided overly optimistic forecasts about structured finance products. […] Most importantly, past downturns in housing prices were primarily regional phenomena—the United States had not experienced a nationwide decline in housing prices in the period following WWII." Apparently, risk modeling failed because historical data lead to a wrong assessment of future risks. Since it is common to rely on historical data for risk modeling, one could fundamentally question its usefulness. An extremely negative perspective on risk modeling is held by Taleb (2007) in his bestselling book "The Black Swan".

To assess the validity of such a view, I apply standard time series models to aggregate US housing prices. I show that risk forecasts derived from such an analysis should have made financial institutions very careful. Forecasts made in 2005 or earlier would have produced worst-case scenarios that are worse than the subsequent, actual change in housing prices. The probability of these scenarios is in the range that banks claim to consider in their risk management. Hence, traditional risk models are not to blame. It seems instead that risk managers failed to apply their toolboxes correctly, or that bank managers failed to respond appropriately to their risk managers' assessments.

There is already a large body of papers which examine the factors that led to the crisis, e.g., Brunnermeier (2008) and Gorton (2008). Many of these papers focus on agency problems, transparency and structural breaks. According to Demyanyk and van Hemert (2008), loan quality deteriorated long before 2007, but the continued surge in housing prices masked the problems because it led to low delinquency rates. Though transparency and model stability are certainly

[1] The fact that Spanish banks, which have been subject to tight regulation, have performed relatively well despite a domestic housing boom is a case in point.

important issues, the present paper shows that a macro-level stress scenario analysis, which does not require intimate knowledge of mortgage markets and structured products, would have produced the right signals.

My use of time series models for forecasting housing prices is motivated by papers which do likewise (e.g., Crawford and Fratantoni, 2003 and Guirguis, Giannikos and Anderson, 2005) and by the fact that risk management models of financial institutions are usually built on a univariate analysis of risk factors. Risk factor scenarios (which take into account the correlation of factors) are then translated into loss scenarios by estimating the factor sensitivities of a bank's positions. This also explains why I do not consider forecasts derived from a fundamental modeling of housing prices. Papers in this branch of the literature often focus on the question of whether a housing market is overvalued (for a survey cf. Himmelberg, Mayer and Sinai, 2005). They focus on current valuation levels and their consequences for *expected* future housing price changes, not on the *variance* in future housing prices. The risk of a housing price decline may be high even though a concurrent fundamental analysis concludes that the housing market is properly valued.

The remainder of the paper is structured as follows. In Section 2, I examine housing price forecasts produced by an AR(1), a first-order autoregressive model. Section 3 shows that the results are insensitive to model specifications. Section 4 concludes.

2 Forecasting housing prices with a simple AR(1) model

What could financial institutions have learned from applying standard risk management tools to readily available historical housing price data? I choose to perform the exercise using data ending in June 2005. The motivation for this is that in August of that year, Alan Greenspan (2005) warned of lower asset values and lower liquidity. After his speech, at the latest, even bank managers who did not care about housing prices before should have procured a housing price scenario from their risk management unit. Note, too, that June 2005 precedes the housing price peak, which was observed in 2006.

Let's assume that you were a risk manager at that time. Your first pick might have been the Case/Shiller national home price index. Quarterly data are available from Q3 1987.[2] Regarding

[2] http://www2.standardandpoors.com.

the model, your first try might have been a first-order autoregressive model, AR(1) for short, which describes the evolution of a variable y_t as follows:

$$y_t = a + \rho y_{t-1} + \varepsilon_t \qquad (1)$$

where ε_t denotes the innovation, a a constant and ρ the autoregressive coefficient. When I fit such a model to quarterly percentage changes of the seasonally adjusted Case/Shiller National home price index, I get a highly significant estimate for ρ (0.918, t-stat= 20.83).

With estimates of the model's parameters at hand, it is easy to generate scenarios far into the future. Based on a distributional assumption for the innovations—I start with the assumption that they are normally distributed—one can generate scenarios based on equation (1). I use a Monte Carlo simulation to generate 100,000 independent paths for quarterly price changes from the third quarter of 2005 to the third quarter of 2008.[3] The simulated quarterly price changes are then used to construct a scenario for the index level, with the Q2 2005 level as the starting value.

Figure 1 depicts the actual index together with two predictions from extreme scenarios: one is the 1% quantile, i.e., in each quarter I select the scenario which is better than just 1% of all scenarios made for that quarter. The other is the 0.1% quantile. Note that in bank risk management, it is common to examine 0.1% quantiles or even more extreme ones. The Basel II regulatory framework, for example, is based on the 0.1% quantile of the annual credit loss distribution.

The scenarios suggest that there was no reason to worry about a decline in housing prices. Even very pessimistic scenarios do not include a sizeable drop in housing prices. The 0.1% worst case shows only a flattening of housing price growth, and small declines three years after the start of the prediction horizon.

But wait a minute. We have fewer than 20 years of data, which are dominated by a long housing price surge. Perhaps the data do not represent housing price dynamics. One check is to run a test for stationarity. Effectively, what we test is whether the variable that we model—quarterly housing price changes—reverts to some normal level, or can potentially explode. A standard Dickey-Fuller (DF) test[4] yields a statistic of -1.36. The null hypothesis of non-stationarity

[3] I use a Monte Carlo simulation because later I examine the likelihood of events that cannot easily be determined with analytically derived confidence intervals.
[4] Results do not depend on the choice of lag length for the DF test.

therefore cannot be rejected. This means that we cannot rule out that housing prices will continue to grow at ever higher rates. Though explosive growth rates mirror the "bubble mentality" that apparently was common in the US, economic reasoning should lead to concerns about the robustness of results. We should not fully trust a model that attaches non-zero probabilities to housing prices that explode and decouple from the rest of the economy.

What would a vigilant risk manager do in such a situation? (i) Try real growth rates.[5] This does not change the picture; we still cannot reject non-stationarity. (ii) Look for a longer time series.

Another widely used housing price index is the HPI, compiled by the Office of Federal Housing Enterprise Oversight (OFHEO).[6] It goes back to 1975. Like the Case-Shiller Index, the HPI is based on the repeat-sales method.[7] I examine real housing price changes to account for secular changes in inflation levels observed during this period, and use the automatic XT-12 ARIMA procedure of the US Census Bureau for seasonal adjustment. Descriptive statistics for the two indexes are shown in Table 1.

Using quarterly data ending in the second quarter of 2005, I perform the Dickey-Fuller test with the HPI and now find that the test statistic is -5.8, which means that non-stationarity can safely be rejected. Fitting an AR(1) model to data ending in Q2 2005 yields a ρ coefficient of 0.527 (t-statistic: 7.69). I then conduct the same scenario analysis as above. The scenarios, shown in Figure 2, look much grimmer than in the analysis with the Case/Shiller index. Both the 0.1% and 1% scenarios stay below the actual index.

What should a financial institution have done with such forecasts? On a one-year horizon, banks typically aim to withstand a worst case that is expected to happen with a probability of 0.1% or less. This roughly corresponds to the annual default probability of a company with a credit rating of A. On a three-year horizon, the cumulative historical default frequency of A-rated companies as reported by Standard and Poor’s (2005) is 0.24%. Since the 2008 housing price level was predicted to occur with a probability of more than 1%, banks conducting the analysis that led to Figure 2 should have prepared to withstand a housing price level that is lower than the one from the third quarter of 2008.

---

[5] I deflate prices with the all-item consumer price index for all urban consumers, available at http://research.stlouisfed.org.

[6] The data are available through www.ofheo.gov.

[7] For information about the differences between the two indexes cf. Leventis (2007).

Assessing a scenario's consequences for a financial institution requires additional steps. A bank has to gauge how such a change will affect the value of its existing assets, its future business and overall economic conditions. Although this is a complex task, it is what risk management units are designed to do: explore a factor movement's consequences on a bank's risk position. Performing such an analysis is beyond this paper's scope. It shall suffice to provide indirect evidence that it should have been obvious in 2005 that a housing price decline such as the one that the model predicted as a worst case would have dramatic consequences.

First, previous banking crises often were associated with real estate crashes. Reinhart and Rogoff (2008) show that the US housing price boom was more pronounced than those that preceded the five big banking crises of the 1970-1990s (Spain, Norway, Finland, Sweden and Japan).

Second, analysts underestimated the probability of a housing crash but not its effects. Analysts of the rating agency Fitch admitted that their models "would break down completely" (Rodriguez, 2007) if home prices declined for an extended period of time. Gerardi, Lehnert, Sherlund and Willen (2008) examine analyst reports from the time before the crash and conclude: "When they did consider scenarios with house price declines, market participants, in the main, appear to have correctly identified the subsequent losses. However, such scenarios were labelled as `meltdowns' and ascribed a low probability" (p. 37).

Third, the scenarios derived here are much more extreme than the one used by Fannie Mae, the government sponsored mortgage financer that collapsed in 2008. In its 2004 annual report, Fannie Mae describes its housing price stress scenario as follows:

> "We develop a baseline scenario that estimates the present value of future credit losses over a ten-year period. We then calculate the present value of credit losses assuming an immediate 5% decline in the value of single-family properties securing mortgage loans we own or that back Fannie Mae MBS. Following this decline, we assume home prices will follow a statistically derived long-term path." (p. 152)

To measure housing price performance, Fannie Mae used the OFHEO housing price index used in this paper. To compare the Fannie Mae scenario to the ones derived with an AR(1), I assume an immediate 8% drop in housing prices because the Fannie Mae scenario is done for nominal prices and because inflation hovered around 3% in the four quarters ending Q2 2005. I further assume that the "statistically derived long-term path" is based on the average growth rate from Q1 1975 to Q2 2005.

Figure 3 compares the Fannie Mae stress scenario to the 0.1%, 1% and 3% worst-case scenarios from an AR(1). An immediate 5% decline may appear to be very extreme. Together with the assumed return to a long-term path, this scenario is rather mild, though. After three years, the Fannie Mae scenario is already better than the 3% worst case derived from time series analysis.

Fannie Mae's objective was "to sufficiently capitalize and hedge our mortgage portfolio and credit guaranty business so that each is able to withstand internal or external 'stress tests' set to at least AA/Aa standard" (Annual Report 2003, p. 9). On a three-year horizon, the historical default frequency of AA- rated issuers has been less than 0.1% (cf. Standard & Poor's, 2005). If this paper's AR1(1) analysis is reliable, the Fannie Mae scenario was much too benign. On a three-year horizon, it should be close to the 0.1% worst case; in fact, it ends up above the 3% worst case. Instead of securing a AA rating, the Fannie Mae scenario was sufficient to guarantee a rating of only BB+, which is associated with a three-year default probability of 3.6%

Of course, one could argue that financial institutions would have suffered less than they actually did if prices had started to fall in 2005, as in the worst-case scenarios in Figure 2. Banks likely would have raised their lending standards as early as 2006 and set aside more capital. Perhaps the crisis occurred only because the boom continued well into 2006. I therefore examine the ex ante probability that prices evolved like they actually did. I find that a fairly narrow band around the actual path of the housing price index is sufficient to generate probabilities that are large enough to be considered by banks in their stress scenarios. Take a band that is +/-2% in the first quarter for which a prediction is made; its width then linearly increases to +/-3.5%. The band is shown in Figure 4. Again, I perform 100,000 trials. Now I count how many of the simulated paths stay within the band throughout Q3 2005 to Q3 2008. The simulated probability of being in the band over this entire period is 0.25%. Recall that this is the three-year default probability associated with a credit rating of A. Therefore, the actual development of housing prices was well within the realm of possibilities that financial institutions should consider in their risk management analysis.

3 Robustness checks

Prediction horizon

In the previous section, I started the analysis in the second quarter of 2005. Conclusions do not change when the exercise is started earlier. The earlier the starting date, the lower the initial index level and the higher the probability that the scenario for 2008 is below the actual value. It is therefore more interesting to examine the expected probability that the index would follow the path that it actually took. If the analysis is started in Q4 2003, the forecast horizon is five years. The five-year cumulative default frequency of A-rated bonds is 0.6%. The scenario analysis shows that a fairly narrow band is sufficient to guarantee that the probability of staying within the band over the entire period from 2003 to 2008 is also 0.6%; specifically, I consider a band that is +/-3.5% in Q1 2004 and whose width then linearly increases to +/-6%. This is still so close to the actual development that, already in 2003, financial institutions should have prepared themselves for a scenario such as the one that eventually materialized.

Model specification

I have selected an AR(1) because it is a very simple specification and yet captures important aspects of housing price dynamics. Also, I assumed normally distributed innovations, again because it is the default assumption.

Let us examine these two assumptions in turn. Model selection often is based on the Bayes information criterion (BIC) or the Akaike information criterion (AIC).[8] I now consider all combinations of AR(p) and MA(q) processes, where p and q vary independently from 0 to 4. This gives a total of 25 different models. Additionally, each of these models is also estimated with a GARCH(1,1) specification, producing another 25 models. The fitted models are therefore taken from the following set of specifications:

$$\begin{aligned} y_t = a + \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-q} + \varepsilon_t, \qquad p \in \{0,1,\ldots,4\},\ q \in \{0,1,\ldots,4\}, \\ \sigma_t^2 = \sigma^2 \ \ or \ \ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \alpha_2 \sigma_{t-1}^2 \end{aligned} \tag{2}$$

[8] The BIC is defined as BIC = $k \ln(n) - 2\ln(L)$, and the AIC as AIC = $2k - 2\ln(L)$ where k is the number of estimated parameters, n is the number of observations and L is the likelihood.

Using data ending in the second quarter of 2005, the best BIC (-807.2) is achieved by an ARMA(3,3) with GARCH(1,1); and the best AIC (-837.8) by an ARMA(3,4) with GARCH(1,1). If the ARMA(3,3) with GARCH(1,1) specification is used to predict scenarios from Q2 2005 to Q3 2008, the worst-case scenarios are much more extreme than with the AR(1) specification. The 1% worst-case for the Q3 2008 index level is 14% below the realized value; with an AR(1), the difference is only 2%.

One could think of other approaches to modeling housing price dynamics. Crawford and Fratantoni (2003) examine regime-switching models on regional US housing prices. They conclude that, out-of-sample, simple ARIMA models perform just as well or better. Note too, that this paper's conclusions would be questioned only if a more sophisticated model produced risk forecasts that were much more benign than the ones obtained here.

Distributional assumptions and estimation error

So far I have assumed normally distributed errors when generating the scenarios. A careful risk manager might question the adequacy of this assumption. Indeed, the Shapiro-Wilk test rejects the hypothesis that residuals from an AR(1) model estimated with data ending in Q2 2005 are normally distributed at a significance level of 0.29%.

A common procedure for dealing with non-normal innovations is the bootstrap, which often is combined with procedures that account for parameter uncertainty. In the following, I implement the bootstrap procedure suggested by Pascual, Romo and Ruoz (2001), which is designed to deal with both parameter uncertainty and non-normal innovations. For the AR(1) that I examine, the bootstrap's structure is as follows:

(1) Estimate $y_t = a + \rho y_{t-1} + \varepsilon_t$ with data until T = Q2 2005. Use the estimated parameters $\hat{a}$ and $\hat{\rho}$ to determine the residuals $\hat{\varepsilon}_t$ (y_0 is herein assumed to be zero).

(2) Construct a bootstrap replicate according to $y_t^* = \hat{a} + \hat{\rho} y_{t-1}^* + \omega_t^*$, t = 1,…,T, where $y_t^* = y_t$ for t = 1 and ω_t^* is a random draw from the residuals $\hat{\varepsilon}$ determined in step (1).

(3) Estimate an AR(1) on the bootstrap replicate from step (2). Denote the estimates by $\hat{a}^*$ and $\hat{\rho}^*$.

(4) Generate a *k*-step ahead scenario according to $y^*_{T+k} = \hat{a}^* + \hat{\rho}^* y^*_{T+k-1} + \omega^*_{T+k}$, where the ω^*_{T+k} are again drawn from the residuals $\hat{\varepsilon}$ determined in step (1).

(5) Repeat steps (2) to (4) 100,000 times. Determine the quantiles for each *k*.

Applying this approach to the OFHEO housing price index produces a 1% worst case for the third quarter of 2008 that is 3.2% below the realized index level—very close to the 2% difference in the base case with normally distributed innovations.

Logarithmic changes

I checked whether using logarithmic index changes instead of discrete ones changes the picture. It does not, which is not surprising given that quarterly housing price changes are in the narrow interval [-0.025, 0.328]. With the AR(1) specification, the 1% worst case for Q3 2008 is 2% below the actual value—the same difference that is obtained with simple percentage housing price changes.

Seasonal adjustment

I used the original series instead of the seasonally adjusted one. Differences are minor. In both cases, an AR(1) leads to a 1% worst case for Q3 2008 that is 2% below the actual index level.

Nominal price changes

So far I have predicted real housing price changes for the OFHEO index. An examination of nominal housing price changes does not lead to a change in conclusions. The 1% worst case is now 0.3% above the Q3 2008 level. The Fannie Mae scenario examined in the previous section is again above the 3% worst-case for Q3 2008. Also note that an alternative way of producing nominal price scenarios is to analyze real price changes and combine the resulting predictions with an assumed inflation rate. In the third quarter of 2008, the mean five-year inflation rate expected by participants in the Survey of Professional Forecasters was 2.65%.[9] When I inflate the real worst-case scenarios from section 2 accordingly, the 1% worst case for Q3 2008 is 6% below the actual nominal index value. According to this approach, nominal worst case scenarios are therefore more extreme than real ones.

4 Conclusion

Since the 1990s, financial institutions have put significant efforts into internal risk management systems. Stress testing is an important risk management element. Bear Stearns, the investment bank that collapsed in 2008, writes in each of its 2004-2006 annual reports, "Stress testing (also referred to as scenario analysis) measures the risk of loss over a variety of extreme market conditions that are defined in advance. Stress testing is a key methodology used in the management of market risk as well as counterparty credit risk."

In this paper, I have used standard time series models to derive stress scenarios for the US housing market. I consider worst-case scenarios whose probability of occurring conforms to the default probability typically targeted by banks. Predictions made in 2005 or earlier would not only include housing price levels that are lower than the actual ones at the end of 2008, they also would include scenarios in which housing prices continue to increase until 2006 and then fall sharply—much as they did in reality.

Therefore, financial institutions conducting stress scenarios in 2005 or earlier should have factored in a housing crash such as the one that actually occurred from 2007 on. As a consequence, they should have set aside more capital, and followed a more conservative lending and investment policy, than they did. This could have prevented the crisis. In any case, it should have reduced its severity.

These observations help to answer the question of what went wrong in the years before the crisis. Some observers point to fundamental deficiencies in standard risk modeling: distributional assumptions often do not include fat tails; reliance on historical data runs the risk of missing extreme events that have not occurred in the past; responsiveness to structural breaks is often

[9] See http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/.

slow; modelers might have an incentive to pick assumptions that suit their interest. Without doubt, many models suffer from such problems on the micro-level. This does not explain, however, why financial institutions did not produce correct macro-scenarios. I use standard models, which should be bread and butter to any risk manager. These models assume that innovations are normally distributed. Robustness checks show that more sophisticated modeling would have led to more extreme scenarios.

To sum up, we should be careful about concluding that shortcomings in the traditional risk modeling approach played an important role in creating the recent financial crisis. It was not the models that failed to warn of a housing crash. Either risk managers failed to use their models in the way that is taught in a basic risk management course, or bank managers failed to respond to the signals their risk management units produced. Such insights are important when searching for ways to prevent future crises.

References

Brunnermeier, Markus K. 2009. “Deciphering the liquidity and credit crunch 2007-08.” Journal of Economic Perspectives, forthcoming.

Crawford, Gordon W., and Michael C. Fratantoni. 2003. “Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices.” Real Estate Economics, 31: 223-243.

Demyanyk, Yuliya, and Otto Van Hemert. 2008. “Understanding the subprime mortgage crisis.” http://ssrn.com/abstract=1020396.

Gerardi, Kristopher, Andreas Lehnert, Shane M. Sherlund, and Paul Willen. 2008. “Making sense of the subprime crisis.” Brookings Papers on Economic Activity, forthcoming.

Gorton, Gary B. 2008. “The subprime panic.” Yale ICF Working Paper No. 08-25.

Greenspan, Alan. 2005. “Reflections on central banking.” Speech given at a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, August 26.

Guirguis, Hany S., Christos I. Giannikos, and Randi I. Anderson. 2005. “The US housing market: asset pricing forecasts using time varying coefficients.” The Journal of Real Estate Finance and Economics, 30: 33–53.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. “Assessing high house prices: bubbles, fundamentals and misperceptions.” Journal of Economic Perspectives, 19: 67–92.

Leventis, Andrew. 2007. “A note on the differences between the OFHEO and S&P/Case-Shiller house price indexes.” OFHEO.

Pascual, Lorenzo, Juan Romo, and Esther Ruiz. 2001. “Effects of parameter estimation on prediction densities: a bootstrap approach.” International Journal of Forecasting, 17: 83-103.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2008. “Is the 2007 U.S. sub-prime financial crisis so different?” NBER Working Paper No. W13761.

Rodriguez, Richard. 2007. “Absence of fear.” Speech given at the CFA Society of Chicago, June 28.

Standard and Poor's. 2005. Annual global corporate default study: Corporate defaults poised to rise in 2005. Standard and Poor's.

Taleb, Nassim Nicholas. 2007. The black swan. New York: Random House.

Table 1: Descriptive statistics for the Case/Shiller National Home Price Index and OFHEO House Price Index (HPI)

Analysis is based on real, seasonally adjusted quarterly percentage changes. The HPI data starts in Q1 1975, the Case/Shiller data in Q1 1987.

	Mean	Standard deviation	Minimum	Maximum
Index start date – Q2 2005				
HPI	0.36%	1.04%	-2.54%	3.28%
Case/Shiller	0.68%	1.27%	-2.66%	3.60%
Index start date – Q3 2008				
HPI	0.29%	1.12%	-3.26%	3.28%
Case/Shiller	0.26%	1.89%	-6.89%	3.60%

Figure 1: Case/Shiller National Home Price Index and worst-case scenarios based on an AR(1) model estimated in August 2005

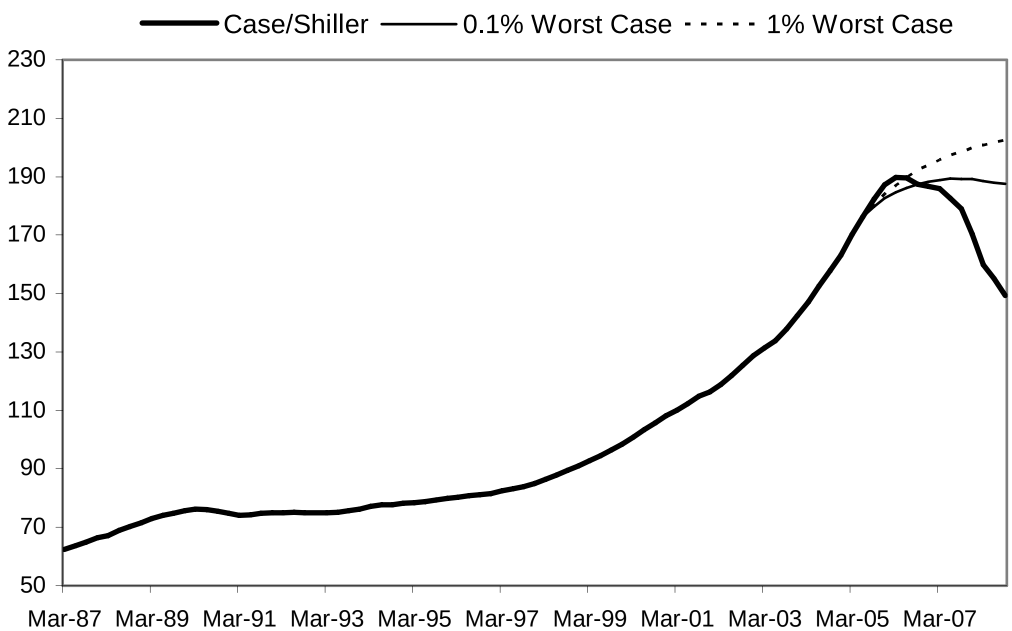

Figure 2: OFHEO House Price Index (HPI) and worst-case scenarios based on an AR(1) model estimated in August 2005 (deflated with CPI)

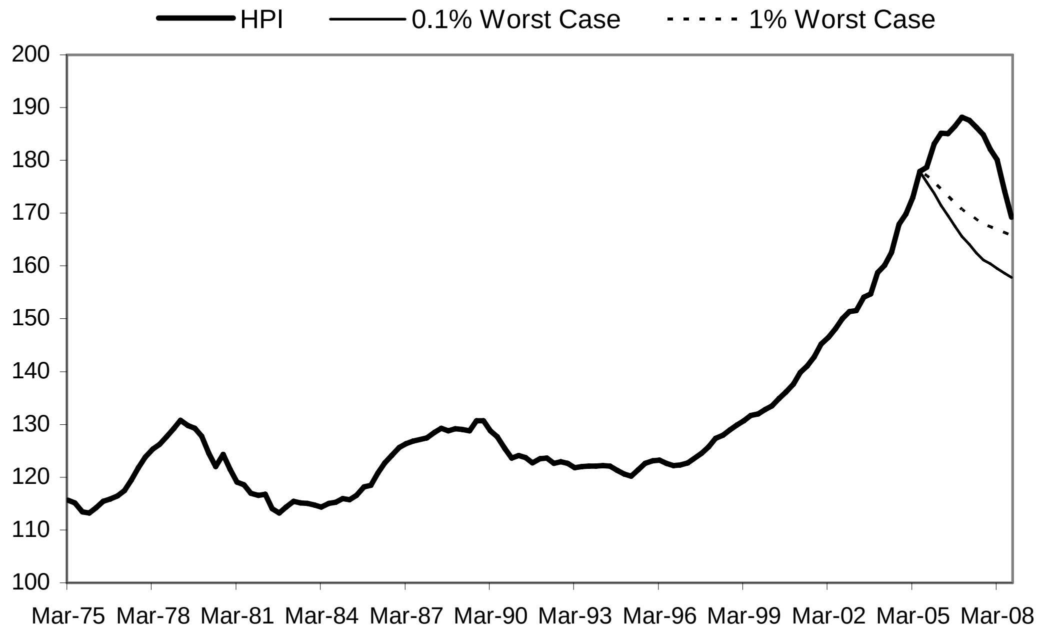

Figure 3: OFHEO House Price Index (HPI), Fannie Mae stress scenario and worst-case scenarios based on an AR(1) model estimated in August 2005 (deflated with CPI)

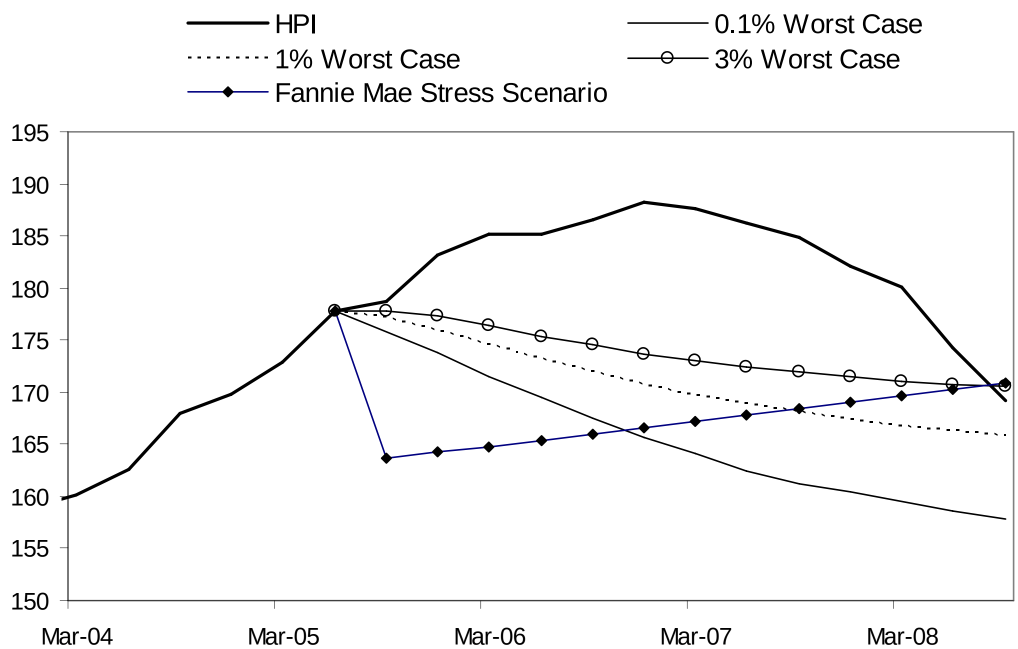

Figure 4: What was the probability of a first-up-then-down scenario for the OFHEO House Price Index HPI (as of August 2005)?

The simulated probability of staying within the band around the actual index throughout Q3 2005 to Q3 2008 is 0.25%.