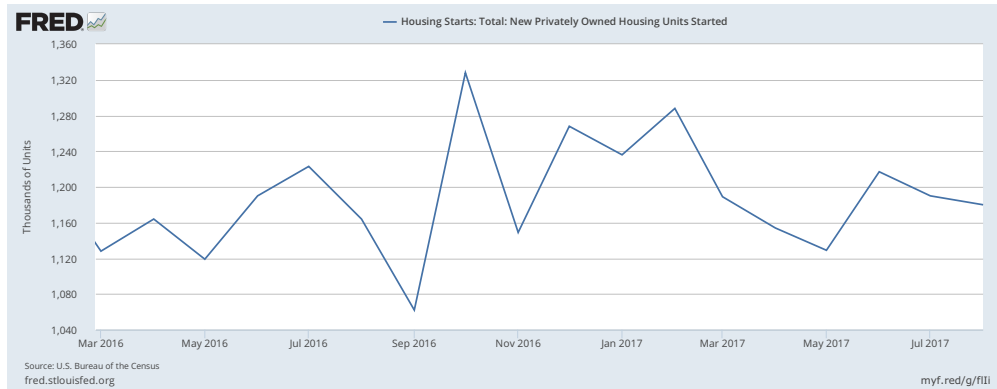# An Example VAR[1]

To illustrate both VARs and the Stata coding involved, we will explore a forecasting issue in macro economics. The background is the decline in the U.S. housing sector since Spring 2017. To motivate the issue, note that housing starts, historically a very good leading economic indicator, had fallen for 5 of 6 months through August 2017.

This has led some to argue that the risk of a U.S. economic downturn within the next 2 years has risen. More recently, existing housing sales prices showed a decline over the summer 2017 as well.


Those arguing that we shouldn't worry too much about bad housing data have raised two points, first that because the slump is due to tight supply (likely related to higher construction costs), it isn't a problem. Second, because homebuilders themselves are optimistic, the bad data are just temporary and housing will soon recover. I am not too interested in the former

---

[1]These are undergraduate lecture notes. They do not represent academic work. Expect typos, sloppy formatting, and occasional (possibly stupefying) errors.

because when housing starts have slumped before (usually leading into recession) supply is usually tight. To me, this feels like a "this time is different" argument. The second though is reasonable. Homebuilders might have information that is not included in other variables that predicts future housing trends.

To capture homebuilder sentiment, I use the National Association of Homebuilders Survey of its members sentiment. Note that it closely tracks housing prices.
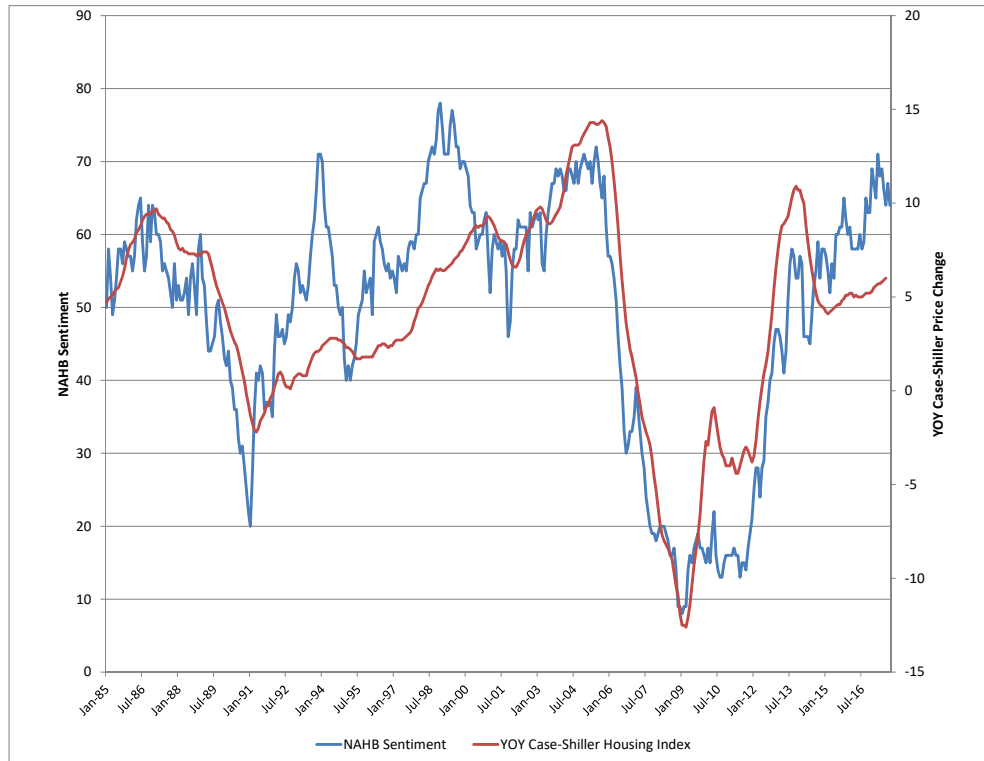
Ideally, I would include this time-series directly into a model that estimates recessionary risk. But because it begins only in January 1985, I would lose too many recessions. So I will instead develop a forecasting model of housing starts and prices that includes homebuilders' sentiment. If optimism leads to much stronger predictions about upcoming starts or prices, then there may be something to the argument that optimism should temper our concerns about disappponting housing data. We will construct a VAR or VECM to perform this forecasting exercise.

We pick up this project with a system that includes the first differences of population adjusted housing starts, U-3 unemployment, and the month-to-month annualized change in the Case-Shiller housing price index. Note that I am initially ignoring the supply and sentiment

factors. i will then add in the monthly supply of houses on the market (to get at tight supply) and the NAHB sentiment variable to see if they matter for forecasting.

You should have access to my *do* and *log* files that show the results that follow. Here, I will just outline the same steps that they show:

1. The *varsoc* command is used to pick lag length.

2. The *var* command then conducts the estimation. Because we are using the Cholesky Decomposition to identify our IRFs, order matters. I begin my system with unemployment. There are tw reasons why I list it as the most exogenous. First, as an aggregate variable, it is affected by other factors besides housing and thus might respond slowly to shocks in the housing sector. Second, unemployment is known to lag the business cycle (making it a poor leading economic indicator). I list housing starts second. Because starting a house takes planning (and a permit), I assume it cannot respond to shocks to prices in the same month. As an asset price, housing prices are assumed to respond immediately to any other shock.

3. The *irf create* command creates my impulse response functions.

4. The first way that I display my IRFs is the *irf graph* command. Note that I list both *oirf* and *fevd*. *oirf* gives me the orthogonalized (identified using Cholesky) IRFs that I care about. We rarely want to use just *irf*. *fevd* reports the forecast error variance decompositions. This is another common way to report VAR results It calculates for each variable, X periods ahead, what fraction of the forecast errors are explained by each of the three shocks in the model. The effects for each period, must add to one. Variance decompositions are a way of quantifying how important each shocks is in teh data used to fit the model.

5. I then, one at a time (it wouldn't matter if I did them both at once) repeat this process including my supply and sentiment variables. I list sentiment last under the assumption that as an expectation, it can move very quickly.

The most striking result is that the sentiment variable does seem to have predictive power. As builders become more optimistic, housing prices and starts rise. The IRFs here are statistically significant. The homebuilders themselves seem to know something that is not captured in the other variables.

6. Before I settle on this VAR as my forecasting model, I want to test for cointegration using the vecrank command.

Figure 1: Forecasts of Housing Starts with and Without NAHB Sentiment
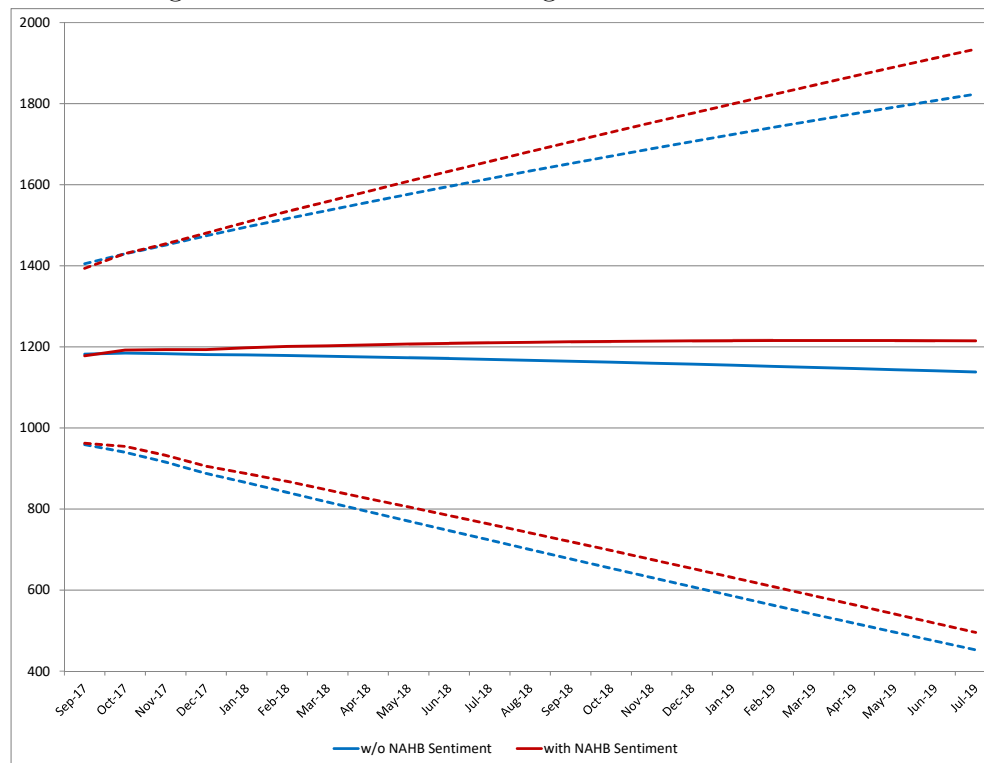
7. Having found evidence of cointegration, I then estimate three vector error correction models using the *vec* command.

8. I then compute my forecasts for each *vec* model using the *fcast compute* command. Note that the default is to forecast only one-period ahead. I thus include the detail *step*(24) to give me forecasts two years ahead.

9. I then display my forecasts using *fcast graph*. I am especially interested in comparing the forecasting model with and without NAHB sentiment. Below, I show results for the forecast of housing starts (the point is the same if I display housing prices instead.)

Based on these results, I argue the following:

a. The homebuilders know something. Including their sentiment switches me from expecting further declines in housing starts to expecting increases. However,

b. The confidence bands are huge in both cases and barely change when I add homebuilders' sentiment. This isn't too surprising. Asset prices are notoriously hard to predict. I thus

4

recommend that agents stay agnostic about the future path of housing data– HAHB sentiment should not be used to rule out future declines and further declines will continue to raise the risk of an economic downturn. That builders are optimistic makes me only feel a little bit better about this summer's crappy housing data.

c. The same thing applies to the forecasts for housing prices.

d. I have made numerous choices when designing this model. These include different orderings, lag lengths, using the *vec* instead of the *var*, etc. I took some time to see if my claims from $a - c$ are robust to other choices by running some alternate models. If the only way I can make these claims to is do the model exactly as I have done here, then I should not have much confidence in my claims. Here, however, alternate models always (in the cases that I tried) deliver the same conclusions.