# Empirical Models

The MW-n-Rents Guys

December 11, 2020

Exogeneity assumption? Something more/less stringent? In this note we run some simulations to figure that out. We run both the static and distributed lag model on several rent variables: some constructed under the strict exogeneity, and some not.

## 1. Models

### Static Model

Static model is

$$\ln y_{it} = \alpha_i + \hat{\delta}_t + \beta \ln \underline{w}_{it} + \epsilon_{it} \tag{1}$$

where:

- $y_{it}$ is rent per square foot for the Zillow SFCC series,
- $\underline{w}$ is the minimum wage,
- $\alpha_i$ and
  $hat\delta_t$ are fixed effects.

To remove the zipcode-specific interecept we estimate the model in first-derences:

$$\Delta \ln y_{it} = \delta_t + \beta \Delta \ln \underline{w}_{it} + \Delta \epsilon_{it} \tag{2}$$

### Distributed lag model

$$\Delta y_{it} = \delta_t + \sum_{r=-s}^{s} \beta_r \Delta \underline{w}_{i(t-r)} + \Delta X_{ct}^{'} \eta + \Delta \varepsilon_{it} \ , \tag{3}$$

where $s$ is windows of leads and lags.

## 2. Simulation

We start by simulating a raw dataframe

```r
library(dplyr)

n_zipcodes = 500  # number of zipcodes
n_periods  = 61   # number of periods (5 years)

df <- data.frame(
  "zipcode" = rep(1:n_zipcodes, each = n_periods),
  "period"  = rep(seq(1, n_periods), times = n_zipcodes)
)
```

Below we simulate auto-correlated shocks, rents and the MW variable. Importantly, note that:

- `shock` is simulated independently, so the strict exogeneity assumption with respect to the MW holds.
- `shock2`, on the other hand, violates strict exogeneity. The reason is that, if last period there was a MW event, we increase the shock by 1%.

```
set.seed(23)
# Simulate auto-correlated shocks
m = 0.5
s = 0.6
rho = 0.6

df$iid.shocks <- rnorm(n = length(df), mean = m, sd = s)
df <- df %>% group_by(zipcode) %>%
  mutate(lagged.iid.shock = lag(iid.shocks),
         shock = rho*lagged.iid.shock + iid.shocks,
         lagged.iid.shock = NULL)

# Simulate time-period effects
other.shocks <- rnorm(n = n_periods)
df <- df %>% group_by(period) %>%
  mutate(period.effect = period/4 + other.shocks[period])

# Simulate MW
mw_change_bounds = c(0.5, 1)

df <- df %>% group_by(zipcode) %>%
  mutate(mw_event    = rbinom(n(), 1, prob = 1.5/n_periods), # 1.5 events per zipcode on average
         d_ln_mw = mw_event*runif(n(), mw_change_bounds[1], mw_change_bounds[2]),
         L_d_ln_mw = lag(d_ln_mw),    # Lags
         L2_d_ln_mw = lag(L_d_ln_mw))

# Simulate weird shocks
df <- df %>% group_by(zipcode) %>%
  mutate(shock2 = ifelse(lag(mw_event) == 1, shock*1.01, shock))
```

Finally, with all the ingredients we simulate several rent variables

```
# Simulate rents
mw_eff1 = 0.03
mw_eff2 = 0.015

df <- df %>%
  mutate(d_ln_rent_1 = period.effect + shock
                     + mw_eff1*d_ln_mw,
         d_ln_rent_2 = period.effect + shock
                     + mw_eff1*d_ln_mw + mw_eff2*L_d_ln_mw,
         d_ln_rent_3 = period.effect + shock2
                     + mw_eff1*d_ln_mw,
         d_ln_rent_4 = period.effect + shock2
                     + mw_eff1*d_ln_mw + mw_eff2*L_d_ln_mw)
```

Now we simulated four variables: `d_ln_rent_1`, `d_ln_rent_2`, `d_ln_rent_3` and `d_ln_rent_4`. The variables have the following characteristics:

- `d_ln_rent_1`: strict exogeneity holds and no dynamic effect of MW

- `d_ln_rent_2`: strict exogeneity holds and 1-period dynamic effect of MW
- `d_ln_rent_3`: strict exogeneity violated and no dynamic effect of MW
- `d_ln_rent_4`: strict exogeneity violated and 1-period dynamic effect of MW

Further, note that the effect of the MW assumed is equivalent to the one we find in the paper.

In the following chunk we estimate models that allow for up to 2 lags. The results are pretty encouraging. The first variable identifies no dynamic effect and a correctly sized effect of MW changes. The second variable identifies both coefficients correctly.

```
library(fixest)

df$period <- as.factor(df$period)

fit1.1 <- feols(d_ln_rent_1 ~ d_ln_mw | period, df)
fit1.2 <- feols(d_ln_rent_1 ~ d_ln_mw + L_d_ln_mw | period, df)
fit1.3 <- feols(d_ln_rent_1 ~ d_ln_mw + L_d_ln_mw + L2_d_ln_mw | period, df)
fit2.1 <- feols(d_ln_rent_2 ~ d_ln_mw | period, df)
fit2.2 <- feols(d_ln_rent_2 ~ d_ln_mw + L_d_ln_mw | period, df)
fit2.3 <- feols(d_ln_rent_2 ~ d_ln_mw + L_d_ln_mw + L2_d_ln_mw | period, df)

etable(fit1.1, fit1.2, fit1.3, fit2.1, fit2.2, fit2.3,
       file = "simulations1.tex", replace = TRUE)
```

| Dependent Variables: | | d_ln_rent_1 | | | d_ln_rent_2 | |
|---|---|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables* | | | | | | |
| d_ln_mw | 0.0280*** | 0.0280*** | 0.0279*** | 0.0281*** | 0.0280*** | 0.0279*** |
| | (0.0035) | (0.0035) | (0.0036) | (0.0035) | (0.0035) | (0.0036) |
| L_d_ln_mw | | 0.0026 | 0.0024 | | 0.0176*** | 0.0174*** |
| | | (0.0035) | (0.0036) | | (0.0035) | (0.0036) |
| L2_d_ln_mw | | | -0.0025 | | | -0.0025 |
| | | | (0.0035) | | | (0.0035) |
| *Fixed-effects* | | | | | | |
| period | Yes | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | |
| Observations | 30,000 | 30,000 | 29,500 | 30,000 | 30,000 | 29,500 |
| $R^2$ | 0.99969 | 0.99969 | 0.99969 | 0.99969 | 0.99969 | 0.99969 |
| Within $R^2$ | 0.00185 | 0.00187 | 0.00186 | 0.00185 | 0.00259 | 0.00257 |

*One-way (period) standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

But don't yell victory yet, Diego! The estimations below use the variables among which strict exogeneity is violated (the MW has an effect on next period shock).

Using `ln_rent_3` we incorrectly find a 2-period effect of a size quite similar to the models above (of course, I tweaked the numbers to find this). We should find nothing, since this variable was constructed without lagged effects of the MW.

Using `ln_rent_4` we find an effect in the second period that is too large. The model thinks the direct effect of the MW is the effect of the MW on rents plues the effect of the MW on the shock.

```
fit1.1 <- feols(d_ln_rent_3 ~ d_ln_mw | period, df)
fit1.2 <- feols(d_ln_rent_3 ~ d_ln_mw + L_d_ln_mw | period, df)
fit1.3 <- feols(d_ln_rent_3 ~ d_ln_mw + L_d_ln_mw + L2_d_ln_mw | period, df)
```

```r
fit2.1 <- feols(d_ln_rent_4 ~ d_ln_mw | period, df)
fit2.2 <- feols(d_ln_rent_4 ~ d_ln_mw + L_d_ln_mw | period, df)
fit2.3 <- feols(d_ln_rent_4 ~ d_ln_mw + L_d_ln_mw + L2_d_ln_mw | period, df)

etable(fit1.1, fit1.2, fit1.3, fit2.1, fit2.2, fit2.3,
       file = "simulations2.tex", replace = TRUE)
```

| Dependent Variables: | | d_ln_rent_3 | | | d_ln_rent_4 | |
|---|---|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
| *Variables* | | | | | | |
| d_ln_mw | 0.0281*** | 0.0280*** | 0.0279*** | 0.0281*** | 0.0280*** | 0.0279*** |
| | (0.0035) | (0.0035) | (0.0036) | (0.0035) | (0.0035) | (0.0036) |
| L_d_ln_mw | | 0.0115*** | 0.0113*** | | 0.0265*** | 0.0263*** |
| | | (0.0035) | (0.0036) | | (0.0035) | (0.0036) |
| L2_d_ln_mw | | | -0.0025 | | | -0.0025 |
| | | | (0.0035) | | | (0.0035) |
| *Fixed-effects* | | | | | | |
| period | Yes | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | |
| Observations | 30,000 | 30,000 | 29,500 | 30,000 | 30,000 | 29,500 |
| $R^2$ | 0.99969 | 0.99969 | 0.99969 | 0.99969 | 0.99969 | 0.99969 |
| Within $R^2$ | 0.00185 | 0.00217 | 0.00215 | 0.00185 | 0.00352 | 0.00348 |

*One-way (period) standard-errors in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

## 3. Takeaways

Given these simulations, I (SH) think our effect is consistent with two scenarios:

- Strict exogeneity holds and MW directly affects rents in same and following months.
- Strict exogeneity doesn't hold. The MW affects rents in same month and shock in the following month, which we see as a 2-period dynamic effect of the mW in our model.