CEP 17-03

Do Real Estate Shocks Affect Corporate Investment?

Timothy Grieder
Carleton University
& Bank of Canada

Hashmat Khan
Carleton University

January 2017; revised 19 December 2017 and 29 December 2017

CARLETON ECONOMIC PAPERS

Carleton UNIVERSITY

Department of Economics

1125 Colonel By Drive
Ottawa, Ontario, Canada
K1S 5B6

Do Real Estate Shocks Affect Corporate Investment?*

Timothy Grieder†
Carleton University &
Bank of Canada

Hashmat Khan‡
Carleton University &
Ottawa-Carleton GSE

December 29, 2017

Abstract

Chaney, Sraer and Thesmar (2012) find that over the 1993-2007 period, a $1 increase in collateral (the value of real estate a firm actually owns) leads the representative US public corporation to raise its investment by $0.06. We successfully replicate their result in a narrow sense. In a wide sense, however, we demonstrate their results are highly sensitive to arbitrary Winsorization thresholds. The extent to which real estate shocks affect corporate investment—the collateral channel—therefore, remains an open question.

JEL classification: D22, G31, R30

Key words: Collateral, Real Estate Prices, Corporate Investment, Winsorization

*We thank Thomas Davidoff for comments.

†Department of Economics, Loeb Building, 1125 Colonel By Drive, Carleton University, Ottawa, K1S 5B6, Canada, and Financial Stability Department, Bank of Canada, 234 Wellington Street, Ottawa, K1A 0G9, Canada, *E-mail:* timothy.grieder@carleton.ca. The views in this paper are those of the author and do not necessarily reflect those of the Bank of Canada.

‡Corresponding Author, Department of Economics, D891 Loeb Building, 1125 Colonel By Drive, Carleton University, Ottawa, K1S 5B6, Canada. *E-mail:* hashmat.khan@carleton.ca, Tel: +1-613-520-2600 (ext 1561), Fax: +1-613-520-3906.

In "The Collateral Channel: How Real Estate Shocks Affect Corporate Investment", Chaney, Sraer and Thesmar (2012)—henceforth, CST—study how changes in the value of real estate a firm owns affects its level of investment. They find that over the 1993-2007 period, a $1 increase in collateral (the value of real estate a firm actually owns) leads the representative US public corporation to raise its investment by 6 cents ($0.06). CST emphasize the economic significance of this sensitivity as real estate represents a sizeable fraction of the tangible assets that firms hold on their balance sheet.

In this note, we first replicate their findings in a narrow sense by combining their submitted data and software provided on the *American Economic Review*'s website with the required Compustat Data. We then show their findings on the collateral channel are highly sensitive in a wide sense. In particular, their findings depend on the choice of data Winsorization threshold prior to the empirical analysis. As the Winsorization threshold increases, the point estimate monotonically increases. When the threshold is 0% (i.e., using the unaltered data), the point estimate is close to zero suggesting that real estate shocks do not affect corporate investment. The point estimate is as high as 6 cents only if the data is Winsorized using a 5% threshold level. Winsorization is, therefore, necessary for evidence favourable to the collateral channel. We discuss potential reasons that produce this sensitivity.

The practice of Winsorization is common and widely used in a variety applied econometric contexts in finance and labour economics studies to 'clean the data'.[1] Winsorizing reduces values at the low and high ends of the sorted data to a pre-determined cut-off point. As discussed in Bollinger and Chandra (2005), however, Winsorizing data can induce or exacerbate biases in estimation. We demonstrate that this problem is severe for CST's finding: their main result (6 cents increase in investment per $1 increase in the value of real estate the firm owns) is obtained *only if* the data is Winsorized. Recently, in the 2017 Presidential Address to the American Financial Association, Harvey (2017) has drawn attention to this form of data exclusion, among other choices made in applied work. He writes: *Researchers make many choices in terms of standardization, log or other transformations, winsorization, and outlier exclusion. If these choices lead to the most significant*

[1]See, for example, Angrist and Krueger (2000), Harvey (2017), among others. Tukey (1962) coined the term Winsorizing or Winsorization in honour of Charles P. Winsor.

results being presented, this is p-hacking. We view our findings as an example of how Winsorization can have serious implications when determining the importance of the collateral channel. The extent to which real estate shocks affect corporate investment, therefore, remains an open question for further research.

I. Replication in a Narrow Sense

In this section we discuss the econometric model and data used in CST and replicate their results in a narrow sense. The main empirical specification from CST (throughout we use same notation as in their paper) is

$$INV_{it}^{l} = \alpha_i + \delta_t + \beta RE\,Value_{it} + \gamma P_t^l + controls_{it} + \epsilon_{it} \tag{1}$$

where INV_{it}^{l} is investment by firm i at time t located in the Metropolitan Statistical Area (MSA) l. Parameters α_i are firm fixed effects and δ_t are time dummy variables designed to capture macroeconomic fluctuations in real estate prices. The variable $RE\,Value_{it}$ measures the market value of real estate assets owned by firm i at time t. Due to data limitations, it is assumed that all real estate owned by a firm is located in MSA l. P_t^l measures real estate prices at time t in area l, and $controls_{it}$ include the amount of cash on the balance sheet, the previous year's ratio of the market-to-book value of equity, and the interaction of real estate prices with controls for ownership of real estate assets (i.e., quintile of firm age, return on assets, firm size measured by the total amount of assets, industry dummy and state dummy).[2] All variables except market value of equity to book value of equity are normalized by past year's Property Plant and Equipment (PPE).

Parameter β in (1) is the coefficient of interest in CST's empirical analysis since it captures the sensitivity of investment to changes in real estate value the firm actually owns.[3]

[2]We use the data and programs available for the CST paper on the *American Economic Review* webpage. In addition, we downloaded firm level financial data from COMPUSTAT via WRDS and followed instructions from the readme file to merge this data with what is available on the *American Economic Review* webpage.

[3]Due to data limitations, CST are only able to measure the real estate *holdings* of a firm in 1993, which we denote by $Q_{i,1993}$. To measure the *value* of real estate holdings, CST inflate $Q_{i,1993}$ by MSA level real estate prices, P_t^l, and then normalize by $PPE_{i,t-1}$. Thus, RE $Value_{it} \equiv \frac{Q_{i,1993*P_t^l}}{PPE_{i,t-1}}$. Therefore, the sensitivity of investment to real estate shocks is given by: $\frac{\partial INV_{it}^l}{\partial P_t^l} = [\beta * \frac{Q_{i,1993}}{PPE_{i,t-1}} + \gamma]$. Since the focus of this note is on replication, we leave the interpretation issue of β aside.

Since real estate prices, and hence the value of real estate owned by a firm, are likely correlated with its investment opportunities, CST use the Instrumental Variables (IV) approach. They use the following first-stage specification

$$P_t^l = \alpha^l + \delta_t + \gamma E^l \times IR_t + u_t^l \tag{2}$$

where α^l are MSA level fixed effects, δ_t are time dummy variables. E^l are the housing supply elasticities developed in Saiz (2010), and IR_t is the nationwide real interest rate at which banks refinance their home loans at time t.[4] CST use two sets of MSA level real estate prices in their analysis: office prices and residential prices. All of our results use residential prices since they are publicly available. We do not have access to the office price index. Nevertheless, as mentioned by CST, results are similar regardless of price index used.

The firm level financial data is obtained from Compustat. CST apply an initial selection criteria to this data:

> *We keep firms whose headquarters are located in the United States and exclude from the sample firms operating in the finance, insurance, real estate, construction, and mining industries, as well as firms involved in a major takeover operation. We require firms to have available data every consecutive year they appear in the sample. We keep only firms that appear at least three consecutive years in the sample.* CST p. 2383-2384.

We do the same and refer to the data obtained after this initial selection criteria as the 'unaltered data' (no Winsorization).

The final step CST perform prior to estimation is to Winsorize the unaltered data. In general, Winsorization replaces extreme values at the low and high ends of the sorted data to arbitrarily chosen cut-off points. For instance, if a 5% threshold is used, the bottom 5% of observations are replaced with the value corresponding to the 5th percentile and those above the 95th percentile are replaced with the value corresponding to the 95th percentile.

[4] Davidoff (2016) shows that the interaction of housing supply elasticities and the nationwide interest rate is not a valid instrument for real estate prices. Nevertheless, in addition to replicating CST's results, the main purpose of our note is to highlight the sensitivity of CST's results to arbitrary Winsorization thresholds. As we show in our note, the Winsorization process biases the unaltered data in favour of finding evidence of the collateral channel. This bias remains regardless of the choice of the instrumental variable used.

We apply the same Winsorization thresholds as in CST:

> *Finally, to ensure that our results are statistically robust, all variables defined as ratios are windsorized using as thresholds the median plus/minus five times the interquartile range..... Our results are unchanged if we use this 5 percent windsorizing methods for all variables (footnote 10).* CST, p. 2386

The second and third columns in Table 1 displays CST's reported results using office prices and residential real estate prices, respectively. Column 4 shows our replication of their residential real estate price result using OLS and column 5 shows the IV results. As evident, both our OLS and IV based replication of the results are very similar to the CST results. A $1 increase in collateral (the value of real estate a firm actually owns) leads the representative US public corporation to raise its investment by approximately 6 cents ($0.06).

Table 1: Replication of cst in a narrow sense

	CST's results (IV) Office prices	CST's results (OLS) Residential Prices	Our results (OLS) Residential prices	Our results (IV) Residential prices
RE Value (MSA Prices), β	**0.06*****	**0.053*****	**0.049*****	**0.056*****
MSA Prices, γ	0.22	-0.58	-0.151	-0.135
Cash	0.026***	0.027***	0.031***	0.031***
Market/Book	0.07***	0.068***	0.059***	0.061***
Init. Controls $\times$ MSA Prices	Yes	Yes	Yes	Yes
Year fixed fffects	Yes	Yes	Yes	Yes
Firm fixed Effects	Yes	Yes	Yes	Yes
Observations	17,586	22,894	21,130	17,796
Adjusted R^2	0.33	0.33	0.33	0.33

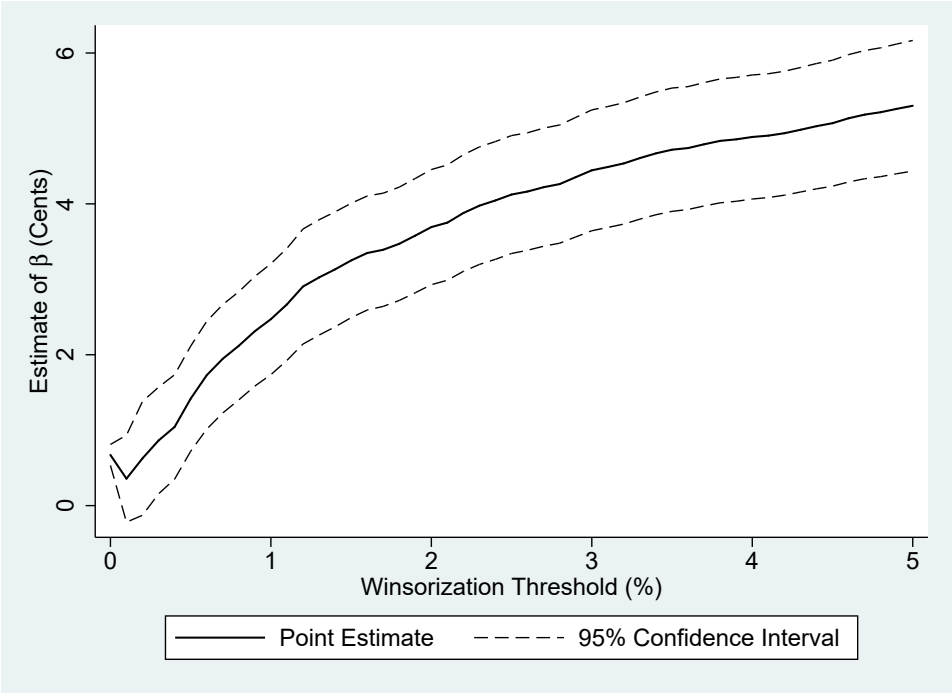
Notes: This table reports results from estimating equation 1. Standard errors are clustered at the MSA-year level. In the IV specifications, standard errors are bootstrapped within MSA-year clusters. *** denotes significance at the 1% level.

II. Replication in a Wide Sense

A. *The Effect of Winsorization*

In this section, we demonstrate that CST's results are quite sensitive to arbitrarily chosen Winsorization thresholds. Figure 1 plots OLS estimates of β from (1) when we apply the Winsorization threshold to the data incrementally increasing from 0% upto a threshold of 5%.[5] As evident, the choice of Winsorization threshold has a large impact on the estimate of β. When the threshold is 0% (i.e., using the unaltered data), the point estimate is close to zero suggesting that real estate shocks do not affect corporate investment. As the Winsorization threshold is increased, the sensitivity of investment to real estate shocks also increases and becomes statistically and economically meaningful.

Figure 1: Effect of Winsorization on estimate of β from equation (1)



As noted in Angrist and Krueger (2000), Winsorizing the original data is performed when it is believed that the extreme values "are exaggerated versions of the true values, but the true

[5]Similar results are obtained if IV is used rather than OLS.

values belong in the tails". In Table 2 we document the smallest and largest extreme values in the unaltered data (no Winsorization). There is little difference between the 1st percentile and the smallest observations for investment and $REValues$. Large differences are evident at the higher end of the distributions. For instance, the 99th percentile for investment takes the value of 3.478, yet the largest observation takes the value 1,226.833. The data for $REValues$ also contain extreme values. While Table 2 shows the smallest and largest extreme values, a general feature of the unaltered data is that there are a continuum of increasing values just beyond the 99th percentile with only a minor number of extreme values far away from the rest of the unaltered data.

Table 2: Summary Statistics for INV_{it} and $REValue_{it}$

Percentile	INV_{it}	$REValue_{it}$	Examples of Extreme Values	INV_{it}	$REValue_{it}$
1	0	0	4 Smallest Observations	-0.288	0
5	0.025	0		-0.088	0
10	0.051	0		-0.058	0
25	0.112	0		-0.055	0
50	0.207	0.202			
75	0.400	1.063	4 Largest Observations	448	37,259
90	0.782	2.460		4,952	64,450
95	1.121	4.226		1,056	66,633
99	3.478	19.721		1,227	169,462

Notes: There are 26,512 INV_{it} observations and 23,687 $REValue_{it}$ observations.

Estimating (1) including all extreme values (e.g., those above the 95th percentile) may exert a large influence on the parameter estimate, and could be either supportive or go against the maintained hypothesis. Normally, one finds evidence in the unaltered data (no Winsorization) and *then* check robustness to the extreme values (e.g., Kaplan and Zingales (1997)). In the CST case, however, it is the other way round. Estimating (1) without Winsorization goes against the maintained hypothesis and we obtain the close to zero cents result (Figure 1).

Extreme observations can be highly informative about the particular theory or channel that one wishes to understand or find evidence for. So a natural question that arises here is: What do the extreme observations represent? We now turn to investigating this question in greater detail.

B. *Do Extreme Values Represent Firms Without Collateral Constraints?*

An important feature of the unaltered data is that the vast majority of firms with very high investment rates do not own any real estate. Thus, extreme investment values likely represent firms without collateral constraints. Winsorization substantially reduces the influence of these observations which biases the estimation result in favour of finding evidence for the collateral channel. Table 3 confirms this point. The first two columns show the quartile and quartile ranges of the unadjusted $REValue$ for all firms in the dataset. The third and fourth columns show that 92% of firms which have their investment Winsorized by CST (i.e., the high investing firms) fall into the lowest quartile of real estate holders. These high investing firms in the bottom quartile of real estate holding firms own no real estate assets. This finding also provides the reason why the point estimate in Figure 1 increases as the Winsorization threshold increases. Detecting the presence of the collateral channel requires a positive relationship between a firm's real estate values and investment. Firms with very high investment rates and no real estate holdings weaken this link. Because the Winsorization process minimizes the influence of these high-investing and no-real-estate-holdings firms, the point estimate increases.

Table 3: Real Estate holdings of high investing firms

$REValue_{it}$ Quartile	$REValue_{it}$ Range	# of firms with investment Winsorized	% of firms with investment Winsorized
1	0 - 0	679	91.76
2	0.002 - 0.202	5	0.68
3	0.202 - 1.063	10	1.35
4	1.061 - 169,461.600	46	6.22
Total		740	100

C. *Are Extreme Values Data Errors?*

One potential reason for an extreme value in the data could be that it is simply a data error. This is often viewed as a justification for either dropping the extreme value or adopting Winsorization. It is possible to check whether the extreme values are correctly captured by Compustat because publicly traded firms are required to submit their financial statements to the Securities and

Exchange Commission.[6] While a manual verification of all data points is unfeasible, we nevertheless manually check the 10 largest investment observations. We verified that 9 of the 10 largest values are accurately captured in the Compustat data.[7] For the remaining observation, we were able to verify investment but not property plant and equipment due to one missing financial statement. Our data check suggests that the extreme values are correctly captured by Compustat and, as a result, there is no justification for Winsorizing the data on the assumption that extreme values are data errors.

III. Conclusion

We re-examine the finding of CST that over the 1993-2007 period, a $1 increase in collateral (the value of real estate a firm actually owns) leads the representative US public corporation to raise its investment by $0.06. We first replicate their paper in a narrow sense and confirm their finding. We then demonstrate that in a wide sense their findings are highly sensitive to the Winsorization threshold. Since the largest investing firms do not hold real estate, Winsorization mechanically reduces their influence and biases the analysis in favour of finding evidence for the collateral channel. Our findings have three implications for future research. First, the empirical relevance of the collateral channel remains on open question. Second, strategic competition and market structure considerations might be more important drivers of US corporate investment than ownership of real estate collateral. Third, more generally, researchers should discuss how Winsorization alters the original data and whether or not it pushes the data closer to the maintained theory of interest.

[6]Financial statements can be found at the SEC's Electronic Data Gathering, Analysis and Retrieval System (EDGAR) website https://www.sec.gov/edgar.shtml.

[7]The table is available upon request.

References

Angrist, Joshua D. and Alan B. Krueger, "Empirical Strategies in Labor Economics," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: Elsevier Science, 2000.

Bollinger, Christopher R. and Amitabh Chandra, "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labour Economics*, 2005, *23* (2), 235–257.

Chaney, Thomas, David Sraer, and David Thesmar, "The Collateral Channel: How Real Estate Shocks Affect Corporate Investment," *American Economic Review*, October 2012, *102* (6), 2381–2409.

Davidoff, Thomas, "Supply Constraints Are Not Valid Instrumental Variables for Home Prices Because They are Correlated with Many Demand Factors," *Critical Review of Finance*, 2016, *5* (2), 177–206.

Harvey, Campbell R., "Presidential Address: The Scientific Outlook in Financial Economics," *Journal of Finance*, 2017, *72* (4).

Kaplan, Steven N. and Luigi Zingales, "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?," *The Quarterly Journal of Economics*, 1997, *112* (1), 169–215.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, August 2010, *125* (3), 1253–1296.

Tukey, John W., "The Future of Data Analysis," *The Annals of Mathematical Statistics*, March 1962, *33* (1), 1–67.