NBER WORKING PAPER SERIES

REPRICING AVALANCHES IN THE BILLION-PRICES DATA

Laura Leal
Haaris Mateen
Makoto Nirei
José A. Scheinkman

Working Paper 29236
http://www.nber.org/papers/w29236

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

We thank Jean-Philippe Bouchaud for encouraging us to use the Billion-Prices data. The work of L. Leal was supported by NSF DMS-1716673 and ARO W911NF-17-1-0578. Code for replication of results can be found at https://github.com/arula10101/BillionPriceAvalanche. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Repricing Avalanches in the Billion-Prices Data
Laura Leal, Haaris Mateen, Makoto Nirei, and José A. Scheinkman
NBER Working Paper No. 29236
September 2021
JEL No. E31

ABSTRACT

Nirei and Scheinkman (2021) proposed an equilibrium model of price adjustments with menu-costs with a finite number of firms and derived a "reproduction number" for repricing and a limit functional form for the distribution of the number of simultaneously price-adjusting firms. We show that the distribution of price-changes in data from the Billion Prices Project is well fitted by this functional form and exhibits a reproduction number that is close to unity, indicating that complementarity in price-changes plays a major role in repricings.

Laura Leal
Department of Operations Research
and Financial Engineering
Princeton University
Princeton, NJ 08540
lleal@princeton.edu

Haaris Mateen
Columbia University
Department of Economics
hm2692@columbia.edu

Makoto Nirei
University of Tokyo
Department of Economics
Japan
mnirei@gmail.com

José A. Scheinkman
Department of Economics
Columbia University
New York, NY 10027
and NBER
js3317@columbia.edu

Code for replication is available at https://github.com/arula10101/BillionPriceAvalanche

1 Introduction

In this paper we use online-price data from the *Billion Prices Project* founded by Alberto Cavallo and Roberto Rigobon, (http://www.thebillionpricesproject.com), to examine implications from state-dependent pricing models on the distribution of the number of simultaneously repricing firms. State-dependent models of repricing often produce complementarity in repricing.[1] In the case of a continuum of firms, often adopted in this literature, this complementarity affects the stationary rate of firms that adjust prices but with a finite number of firms, this complementarity implies that in response to an idiosyncratic shock to a single firm's cost of adjusting prices (Calvo shock) or to the cost of production (technology), firms that are sufficiently close to a price-adjustment threshold should also adjust prices in the same direction, causing an *avalanche* of repricings.

In Nirei and Scheinkman (2021), henceforth NS, the authors propose a rather conventional state-dependent model, except for considering a large number, n, of firms. Firms face a fixed "menu cost" for price adjustment but may benefit from an idiosyncratic Calvo shock that allows for zero-cost price adjustment. There are no aggregate shocks. To avoid the proliferation of state variables, NS assume that agents in the finite n model use the optimal policy from the limit continuum version. NS show that, in a stationary equilibrium, as $n \to \infty$, the distribution of the number of firms that adjust prices in response to a Calvo shock to the cost of price adjustment of a single firm, conditional on the relative price charged by that firm, converges to a *Generalized Poisson Distribution*, henceforth GPD, a two parameter distribution over the non-negative integers.[2] They also show that the coefficient of dispersion (variance/mean) of the GPD equals $\frac{1}{(1-\theta)^2}$, where $\theta < 1$ equals the limit as $n \to \infty$ of the expectation (under the stationary distribution of the continuum model) of the number of firms that adjust prices, conditional on a firm adjusting its price by paying the menu-cost. This "reproduction number" θ characterizes the complementarity in repricing in the limit continuum model and is, in particular, independent of the firm that initially received the Calvo shock. As $\theta \to 1$ the coefficient of dispersion diverges to ∞.

The application of these results to actual data faces at least three obstacles. First, even if observing the number of products repriced in individual avalanches, a researcher cannot identify

[1] For example, Caplin and Spulber (1987), Dotsey et al. (1999), and Golosov and Lucas (2007).

[2] A random variable with values in the non-negative integers has a GPD with parameters $\theta_0 > 0$ and $\theta > 0$ if for each integer $x \geq 0$, $P_x = \theta_0(\theta_0 + \theta x)^{x-1} e^{-\theta_0 - \theta x}/x!$ (Consul and Famoye, 2006, Chapter 9).

the firm that received the Calvo shock and thus is unable to condition on the relative price charged by that firm. In fact, NS show that the unconditional distribution of avalanches also converges as $n \to \infty$ but to a distribution in the more general class of Lagrangian probability distributions treated in Consul and Famoye (2006). NS also show that the coefficient of dispersion of the limit Lagrangian probability distribution exceeds $\frac{1}{(1-\theta)^2}$. Second, one observes histories of successive price-adjustments, instead of independent avalanches started by Calvo shocks with firms' prices sampled from the stationary distribution. Third, data is only available after aggregation in a finite time interval—daily in the case of the Billion Prices Project—and the number of Calvo shocks in any particular day is not observable. To examine these departures, NS simulate histories of successive price-adjustments using calibrated parameter values. The calibrated model is richer than the original theoretical model, since it allows for idiosyncratic technology shocks.[3] The presence of technology shocks gives rise to the possibility of price cuts, even in an equilibrium with a constant positive inflation rate. NS choose parameter values so that the stationary equilibrium for the limit continuum model reproduces the average inflation observed in the US in 1988–2005, the observed volatility of inflation at this average inflation and targets from Golosov and Lucas (2007) and Nakamura and Steinsson (2008). One target that NS are not able to match is the average size of price cuts estimated by Nakamura and Steinsson (2008) (10.5%). For this reason, and since in each avalanche caused by a Calvo shock all price changes have the same sign, NS concentrate on avalanches of price increases. Using the theoretical formula for θ, NS show that for the calibrated values of the parameters, one obtains a θ surprisingly close to 1, indicating a high degree of complementarity in repricings and predicting a high coefficient of dispersion (variance/mean) of the size of price-increase avalanches. In addition, the distribution of individual avalanches from the simulation is very close to a GPD so that neither the impossibility of controlling for the relative price of the firm that initializes the avalanche nor the lack of independence across avalanches seems to cause the observed distribution to be far from the GPD class.[4]

A simulation in NS shows that the aggregation of the number of repricings at the daily frequency impacts the shape of the distribution for small avalanches but has little effect on the estimated θ

[3]The online appendix to Nirei and Scheinkman (2021) contains the necessary theoretical propositions concerning the generalized model.

[4]The simulation of histories in NS leads to an estimate θ^h that exceeds the θ obtained from the calibrated model. NS document that this is the result of time-variation of the distribution of prices caused by the avalanches. θ^h produces a measure of complementarity of price adjustments that takes into account the dynamics.

when compared to the estimate of θ from individual avalanches. The invariance of θ to aggregation is suggested by the fact that the sum of n independent random variables X^i, each with a GPD with parameters (θ_0^i, θ) is again a GDP with parameters $(\sum_i \theta_0^i, \theta)$ (Consul and Famoye, 2006, Theorem 9.1). Thus the numerical results in NS indicate that the variation on the number of daily Calvo shocks and the dependence of successive avalanches within the same day, while affecting the fit of the GPD in the range of small total avalanches, do not much impact the estimated value of θ.[5]

The theoretical results and simulations in NS thus establish new testable implications for state dependent models and in particular for the effect of complementarity on price adjustments.[6] In this paper we examine daily avalanches of price increases in the Billion-Prices data for series from five countries: Argentina, Brazil, Chile, Colombia, and the US (2 series). Our dataset was first used by Cavallo (2018a). We fit GPD's to the data on positive price-change avalanches using two different methods. First, we use moment estimation formulae provided in Consul and Famoye (2006) to estimate the two parameters of the GPD. Second, we assume that the underlying distribution of the daily data follows a GPD with $\theta \leq 1$ and use non-linear least squares (NLS). We get a very good fit for all the datasets that we examine (see Figure 1).

The fit using the NLS method generally performs substantially better than when using moments—the normalized root mean square with NLS is on average 1/4 of the NRMSE obtained with moment estimation. However the estimated θ's are quite close to each other and close to one. These estimates of θ indicate that complementarity plays a major role in repricing.

For robustness we perform a chi-square test, which fails to reject the null hypothesis that the data on the avalanche sizes is distributed as a GPD. We also report in an online appendix that very similar results hold for avalanches of price cuts, which in particular indicates that complementarity plays an important role in price cuts.

[5] The simulations show that for days in which the total number of repricings is small, on average, a large fraction of the total is generated by the daily Calvo shocks, whereas for days with total repricings that exceed the median, the average contribution of Calvo shocks to the total daily avalanche is negligible. This suggests that θ, which depends on the tail of the distribution of avalanches, should be hardly affected.

[6] In models which include also aggregate shocks, the distribution of avalanches would be determined also by the arrival of aggregate shocks and this "background noise" may affect the functional form of the distribution of avalanches.

2 Literature Review

Early empirical contributions on menu-cost models focus on estimating the average size and frequency of price changes (Bils and Klenow (2004); Nakamura and Steinsson (2008); Klenow and Kryvtsov (2008); Gagnon (2009)). Golosov and Lucas (2007) renewed interest on selection effects of menu-cost pricing models, and led to empirical and theoretical papers that discuss the distribution of price changes (Midrigan (2011); Alvarez and Lippi (2014); Alvarez et al. (2016)). Employing the Billion-Prices data, Cavallo (2018a) found a lack of very small price changes, suggesting a role for menu costs or observation costs, as argued by Alvarez et al. (2016).

Online retailer data for Latin American supermarkets in Brazil, Chile, Colombia, Argentina, and Venezuela has been used in Cavallo (2013). This paper has a very distinct goal from ours, namely to construct inflation indices from prices collected from online retailers and compare them to official inflation estimates. The data we use was previously used in Cavallo (2018a) to study the impact of measurement bias on common price stickiness statistics.

3 Data and Results

3.1 Data description

We obtain online price data from the Billion Prices Project.[7] We use daily price data for five countries—USA, Brazil, Argentina, Chile and Colombia that were previously used in Cavallo (2018a).[8] Each unique product in the dataset carries an individual product ID. We use price statistics at the product level.

The Billion Price Project dataset contains several different price variables. We work with "nsfullprice," which fills missing prices by carrying forward (up to 5 months), and excludes sales prices.[9] We apply the same filters as Cavallo (2018a). We restrict our data to products that have at least 100 days of price observations. Table 1 shows the date intervals, the total number of observations, the number of individual products, the number of product categories, and statistics on price changes such as the proportion of price movements that are positive and the median change

[7] See: Billion Prices Project.

[8] Data can be found at: Harvard Dataverse.

[9] Sales occurrence is defined using the v-shaped sale algorithm described in Nakamura and Steinsson (2008).

in log prices conditional on a price increase.

Table 1 shows considerable variation between the four data series. The US supermarket data has more positive price changes than the US department store data. The size of price increase is also higher in the US supermarket series. On average, countries from Latin America have higher proportions of positive price changes. However, the size of price change is smaller than in the US.

For all estimation exercises, we subtract one from the daily positive price series, since the theoretical results concern avalanches following a Calvo shock. This is only an approximation because we cannot count the actual number of Calvo shocks in a day.

3.2 Estimates for θ

Recall that a r.v. X is said to have a GPD with parameters $\theta_0 > 0$ and $\theta > 0$ if X is non-negative valued and if $P_x(\theta_0, \theta)$ denotes the probability that the r.v. X takes a value x, then

$$P_x(\theta_0, \theta) = \theta_0(\theta_0 + \theta x)^{x-1} e^{-\theta_0 - \theta x}/x!, \quad x = 0, 1, 2, \ldots$$

In our case, the r.v. X is the count of how many products had a positive price change on a given day. The probability mass function (pmf) P_x is the probability of observing a day in which exactly x products experience positive price changes. We utilize two methods to produce estimates of the parameters of the GPD described above. Consul and Famoye (2006) describe a moment estimation method that utilizes the sample mean ($\bar{x}$) and the sample variance (s^2) to determine values of θ and θ_0.

$$\theta = 1 - \sqrt{\frac{\bar{x}}{s^2}} \tag{1}$$

$$\theta_0 = \sqrt{\frac{\bar{x}^3}{s^2}} \tag{2}$$

Table 2 presents moment estimates for θ and θ_0 for the six data series with standard errors based on the formulae provided in Consul and Famoye (2006).

As an alternative we present estimates using non-linear least squares. We compute the *counter-cumulative* distribution—the count of how many days in the data have x or more positive price changes. We then minimize the distance between the log of the counter-cumulative of the data and the log of the counter-cumulative of GPDs with $\theta \leq 1$. The fit is rather good especially taking into consideration that GPDs are defined by only two parameters.

We present point estimates and standard errors in Table 3. Although, as expected, NLS fits better than the estimates using moments (see Table 4), the estimates of the crucial parameter of interest, θ, are quite close.[10] Figure 1 shows the plots for the logarithm of the counter-cumulative function of the data (in red) and the estimated GPD (in blue).

For robustness, we also apply a chi-square goodness-of-fit test to compare observed frequencies in the data to the probability mass function $\hat{P}_x^j$ generated using the parameters fitted by NLS to the counter-cumulative distribution for data set j. Again, due to computational limitations we restrict ourselves to a maximum avalanche size of 142. Since simulations in NS show that GPDs do not fit well in the range where the number of total price-changes in a day is small, we perform a goodness of fit test for days in which the total number of avalanches falls in the interval $[q^j, 142]$, where q^j defines the first tercile of the distribution of the number of daily avalanches in datas series j. The null hypothesis is thus H_0 : data in series j is distributed as GPD for $x \in [q^j, 142]$. Let N^j equal to the number of days in which the number of positive repricings is in the interval $[q^j, 142]$. Let $\hat{Q}_x^j := \frac{\hat{P}_x^j}{\sum_{y=q^j}^{142} \hat{P}_y^j}$, the conditional probability implied by our estimate of $\hat{P}^j$. Our chi-square statistics is defined as:

$$\chi^2 = \sum_{x=q^j}^{142} \frac{(O_x^j - N^j \hat{Q}_x^j)^2}{N^j \hat{Q}_x^j}, \tag{3}$$

where O_x^j is the observed number of days in which exactly x positive repricings occurred in series j. The number of degrees of freedom is $142 - q^j$. Table 5 displays p-values associated with the chi-square tests and shows that we fail to reject the null at the 5% level.

Although NS did not attempt to match the average size of price decreases, they were able to match the frequency of price decreases measured by Nakamura and Steinsson (2008). Table A1 in the online appendix gives estimates for θ that shows that the implied reproduction numbers for avalanches of price cuts are also close to unity. Figure A1 shows that the Billion-Prices data exhibit avalanches of price decreases that also approximate a GPD for values of daily avalanches that are

[10]However the differences in estimates are statistically significant. Part of the difference arises for computational reasons. The GPD requires calculating a term that is proportional to the inverse of the factorial of the avalanche size. We are computationally limited to 142! ∼ 2.6e+245. However all data series have a substantial number of days with a total number of positive price changes that exceeds 142, ranging from 4% for USA-2 to 33% for Brazil. Notice that the discrepancy in the estimates of θ for Brazil is the highest, albeit of only 3%.

not too small. A chi-square test also fails to reject the null that the distribution of the number of avalanches of price cuts is a GPD.

4 Conclusion

Nirei and Scheinkman (2021) proposed an equilibrium model of price adjustments with menu-costs and a finite number n of firms. In the model all price changes must be ignited by a Calvo shock and conditional on $\ell << n$ firms repricing, the probability of another firm repricing increases with ℓ. This complementarity is characterized by a "reproduction number"—the expected number of firms that reprice in response to repricing by a single firm that did not benefit from a Calvo shock, as $n \to \infty$. The reproduction number $\theta < 1$ is a characteristic of the mean field approximation with a continuum of firms. In addition, they show that conditional on the relative price of the firm receiving an initial Calvo shock, the distribution of the number of total repricings converges in distribution, as $n \to \infty$, to a Generalized Poisson Distribution, a two parameter distribution over the non-negative integers. Application of these results to actual data faces at least three difficulties. First, a researcher cannot identify the firm that received the Calvo shock and thus cannot condition on the relative price charged by that firm. Second, one observes histories of price adjustments, instead of independent avalanches of price changes, started by Calvo shocks. Third, data is only available after aggregation in a finite time interval—daily in the case of the Billion Prices Project—and the number of Calvo shocks in any particular day is not observable. Simulations in *Nirei and Scheinkman (2021)* with parameter values to reproduce the US experience in 1988-2005 show that the first problem causes minor departures from a GPD. However the use of histories instead of i.i.d. observations yields a larger estimate of θ that also accounts for the dynamics. Aggregation at daily levels produces a distribution that approximates a GPD except when the number of total daily repricings is small, when, on average, Calvo shocks are responsible for a large fraction of the total avalanches. Furthermore, the estimates of θ are hardly affected.

In this paper we examined daily avalanches of price increases in Billion-Prices data that was first used by Cavallo (2018a). We consider six series from five countries: Argentina, Brazil, Chile, Colombia, and the US (2 series). Estimates of θ are consistently close to unity, indicating that complementarity plays an important role in price adjustments and large dispersion (variance/mean) in the number of daily price increases. In addition, the empirical distribution of daily avalanches of price increases is remarkably close to GPDs, especially considering the fact that the GPD is fully determined by two parameters.

References

Alvarez, F., Bihan, H. L., and Lippi, F. (2016). *American Economic Review*, 106:2817–2851.

Alvarez, F. and Lippi, F. (2014). Price setting with menu cost for multiproduct firms. *Econometrica*, 82:89–135.

Alvarez, F. E., Lippi, F., and Paciello, L. (2010). Optimal price setting with observation and menu costs. Working Paper 15852, National Bureau of Economic Research.

Alvarez, F. E., Lippi, F., and Paciello, L. (2011). Optimal price setting with observation and menu costs. *The Quarterly journal of economics*, 126(4):1909–1960.

Bils, M. and Klenow, P. J. (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5):947–985.

Caplin, A. S. and Spulber, D. F. (1987). Menu cost and the neutrality of money. *Quarterly Journal of Economics*, 102:703–726.

Cavallo, A. (2013). Online and official price indexes: Measuring argentina's inflation. *Journal of Monetary Economics*, 60(2):152–165.

Cavallo, A. (2015a). Cavallo (2013) "Online vs Official Price Indexes: Measuring Argentina's Inflation" - Journal of Monetary Economics 60(2), 152-165.

Cavallo, A. (2015b). Cavallo (2018) "Scraped Data and Sticky Prices". Review of Economics and Statistics, Vol. 100, p.105-119.

Cavallo, A. (2018a). Scraped data and sticky prices. *Review of Economics and Statistics*, 100(1):105–119.

Cavallo, A. (2018b). Scraped data and sticky prices. *Review of Economics and Statistics*, 100(1):105–119.

Consul, P. C. and Famoye, F. (2006). *Lagrangian probability distributions*. Springer.

Dotsey, M., King, R. G., and Wolman, A. L. (1999). State-dependent pricing and the general equilibrium dynamics of money and output. *Quarterly Journal of Economics*, 114:655–690.

Gagnon, E. (2009). Price setting during low and high inflation: Evidence from Mexico. *Quarterly Journal of Economics*, 124:1221–1263.

Golosov, M. and Lucas, Jr., R. E. (2007). Menu costs and Phillips curves. *Journal of Political Economy*, 115:171–196.

Klenow, P. J. and Kryvtsov, O. (2008). State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation? *Quarterly Journal of Economics*, 123:863–904.

Midrigan, V. (2011). Menu costs, multiproduct firms, and aggregate fluctuations. *Econometrica*, 79(4):1139–1180.

Mizuno, T., Nirei, M., and Watanabe, T. (2010). Closely Competing Firms and Price Adjustment: Some Findings from an Online Marketplace*. *The Scandinavian Journal of Economics*, 112(4):673–696.

Nakamura, E. and Steinsson, J. (2008). Five Facts about Prices: A Reevaluation of Menu Cost Models*. *The Quarterly Journal of Economics*, 123(4):1415–1464.

Nirei, M. and Scheinkman, J. A. (2021). Repricing avalanches. Working Paper 28654, National Bureau of Economic Research.

United Nations (2018). Un classification of individual consumption according to purpose (coicop).

Table 1. Data overview

	USA-1	USA-2	Argentina	Brazil	Chile	Colombia
Initial date	2008-05-08	2008-03-18	2007-10-07	2007-10-10	2007-10-24	2007-11-13
Final date	2010-07-31	2010-07-31	2010-08-13	2010-08-01	2010-08-13	2010-08-01
Observations	12,037,687	13,554,672	15,019,074	12,401,805	13,651,502	4,737,481
Unique products	21,230	50,004	23,506	19,933	20,600	8,045
Categories	26	32	74	72	72	59
Positive price changes (%)	54.50	34.74	80.69	59.71	58.54	58.02
Median log price increase	0.178	0.117	0.061	0.051	0.075	0.055

USA-1 is US supermarket. USA-2 is US department store. For all other countries the data is on supermarkets. Log price increase is calculated conditional on a positive price change. All prices are non-sale prices for products with at least 100 days of observations.

Table 2. Moment estimation for avalanches of price increases

	Simulation	USA-1	USA-2	Argentina	Brazil	Chile	Colombia
θ	0.895	0.968	0.901	0.964	0.979	0.912	0.880
	(1.1e-06)	(0.0006)	(0.0010)	(0.0005)	(0.0004)	(0.0014)	(0.0014)
θ_0	3.99	4.106	3.447	4.607	3.700	8.402	7.095
	(1.0e-05)	(0.0194)	(0.0104)	(0.0154)	(0.0186)	(0.0175)	(0.0126)

Standard errors are in brackets. All statistics calculated using formulae in Consul and Famoye (2006).

Table 3. NLS estimates for avalanches of price increases

	Simulation	USA-1	USA-2	Argentina	Brazil	Chile	Colombia
θ	0.882	0.959	0.881	0.939	0.944	0.931	0.873
	(1.6e-06)	(2.5e-05)	(4.9e-05)	(5.9e-05)	(9.7e-05)	(7.7e-05)	(4.9e-05)
θ_0	4.019	5.655	3.755	8.665	9.757	7.940	7.771
	(7.4e-05)	(0.001)	(0.002)	(0.004)	(0.006)	(0.005)	(0.003)

Standard errors in brackets.

Table 4. NMRSE Comparison

	USA-1	USA-2	Argentina	Brazil	Chile	Colombia
Moment Estimation	0.219	0.146	0.325	0.455	0.105	0.087
NLS	0.026	0.101	0.049	0.069	0.078	0.098

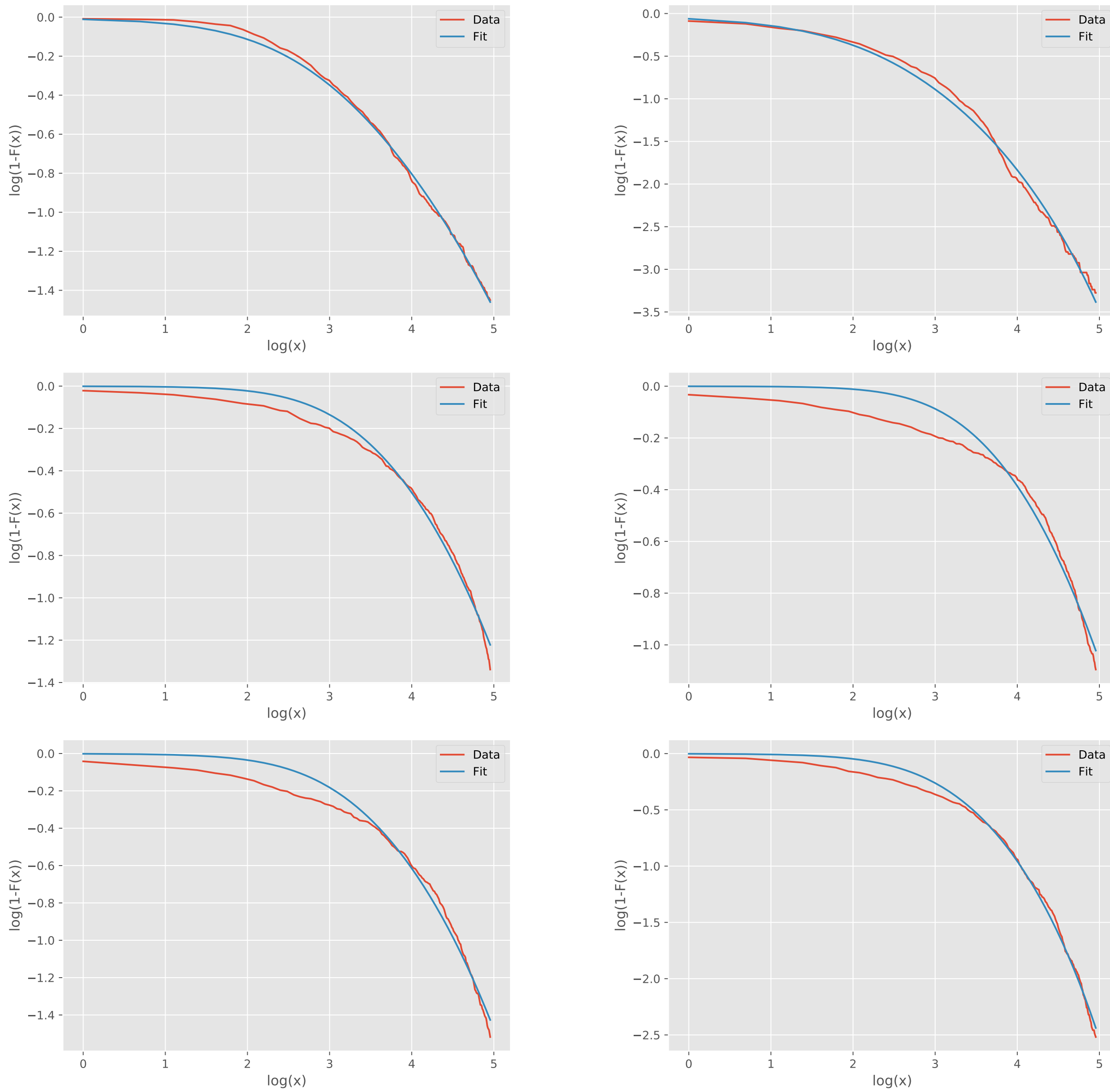
The normalization is the mean.

Table 5. NLS goodness of fit p-values

USA-1	USA-2	Argentina	Brazil	Chile	Colombia
0.783	0.716	0.271	0.157	0.414	0.091

p-values reported from chi-square test. Observations from first tercile of avalanche sizes q^j to 142. q^j values: 25, 9, 44, 61, 36 and 23 respectively.

Figure 1. Log counter-cumulative fit of price increases using NLS



Log counter-cumulative fit—US supermarket (top left), US department store (top right), Argentina supermarket (center left), Brazil supermarket (center right), Chile supermarket (bottom left) and Colombia supermarket (bottom right).

A Online Appendix

Table A1. θ estimates for avalanches of price-cuts

	Moment Estimation	NLS
Simulation $\pi = 0.03$	0.853	0.850
USA-1	0.922	0.960
USA-2	0.924	0.905
Argentina	0.971	0.860
Brazil	0.968	0.935
Chile	0.920	0.891
Colombia	0.854	0.846

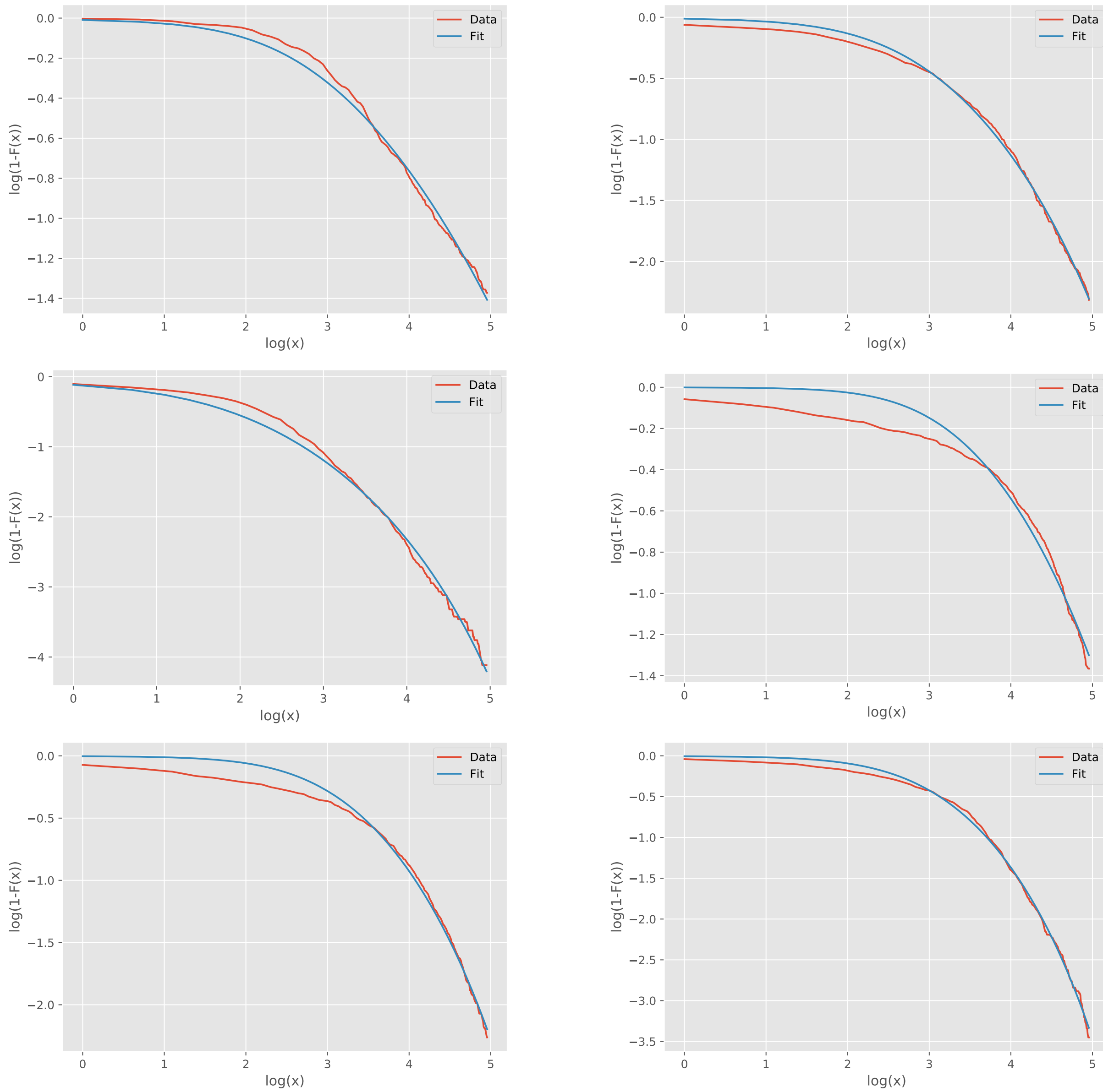
USA-1 is US supermarket. USA-2 is US department store. For all other countries the data is on supermarkets.

Table A2. NLS goodness of fit p-values for price-cuts

USA-1	USA-2	Argentina	Brazil	Chile	Colombia
0.494	0.376	0.255	0.065	0.381	0.059

p-values reported from chi-square test. Observations from first tercile of avalanche sizes to 142.

Figure A1. Log counter-cumulative fit of price-cuts using NLS



Log counter-cumulative fit—US supermarket (top left), US department store (top right), Argentina supermarket (center left), Brazil supermarket (center right), Chile supermarket (bottom left) and Colombia supermarket (bottom right).