FACULTAD ECONOMÍA Y NEGOCIOS UNIVERSIDAD DE CHILE

DEPARTAMENTO DE ECONOMÍA

SERIE DE DOCUMENTOS DE TRABAJO

SDT 412

Missing Aggregate Dynamics: On the Slow Convergence of Lumpy Adjustment Models

Autores:

David Berger

Ricardo Caballero

Eduardo Engel

Santiago, Noviembre de 2015

sdt@econ.uchile.cl

econ.uchile.cl/publicaciones

Missing Aggregate Dynamics: On the Slow Convergence of Lumpy Adjustment Models

David Berger
Northwestern and NBER

Ricardo Caballero
MIT and NBER

Eduardo Engel
U. of Chile and NBER

October 2015[1]

Abstract

When microeconomic adjustment is lumpy, the VAR-estimated persistence of the corresponding aggregated variable is downward biased. The extent of this bias decreases with the level of aggregation, yet convergence is very slow and the bias is likely to be present for sectoral data in general and, in many cases, for fully aggregated data as well. Paradoxically, while idiosyncratic productivity and demand shocks smooth away microeconomic non-convexities and are often used to justify approximating aggregate dynamics with linear models, their presence exacerbates the bias. We propose procedures to correct for the bias and provide various applications. In one of them, we account for the persistence-gap behind Bils and Klenow's (2004) rejection of the Calvo model. In another, we find that the difference in the speed with which inflation responds to sectoral and aggregate shocks (Boivin et al 2009; Mackoviak et al 2009) disappears once we correct for the missing persistence bias.

JEL Codes: C22, C43, D2, E2, E5.

Keywords: Aggregate dynamics, persistence, lumpy adjustment, idiosyncratic shocks, aggregation, aggregate shocks, sectoral shocks, Calvo model, *Ss* model, inflation, investment, labor demand, sticky prices, biased impulse response functions, vector autoregressions.

[1] We are grateful to Filippo Altissimo, Fernando Alvarez, William Brainard, Jeff Campbell, Larry Christiano, Marty Eichenbaum, Xavier Gabaix, Pablo García, Marc Giannoni, Robert Hall, Fabiano Schiaverdi, Jon Steinsson, Eric Swanson, Yuta Takahashi, Harald Uhlig, Joe Vavra and seminar participants at Chicago Fed, Columbia University University of Chicago, FGV (Rio and Sao Paulo), Humboldt Universität, IEIS Stockholm, MIT, NBER SI (EFCE), PUC (Chile), Universidad de Chile (CEA and FEN), University of Maryland, University of Paris, University of Pennsylvania, Uppsala University, Yale University, NBER EFG Meeting, 2nd ECB/IMOP Workshop on Dynamic Macroeconomics, Hydra and the Central Bank of Chile's "The Micro and the Macro of Price Rigidities" workshop for their comments on an earlier version of this paper. We thank Juan Daniel Díaz for outstanding research assistance. Financial support from NSF is gratefully acknowledged. This paper is an extensively revised version of "Adjustment is Much Slower than You Think," NBER WP #9898.

1 Introduction

The dynamic response of aggregate variables to shocks is one of the central concerns of applied macroeconomics. The main procedure used to measure these dynamics consists in estimating a vector autoregression (VAR). In non- or semi-structural approaches, the characterization of dynamics stops there. In other, more structural approaches, researchers wish to uncover underlying parameters from the estimated VAR and use the implied response to shocks as the benchmark against which the success of the calibration exercise, and the need for further theorizing, is assessed.

The main point of this paper is that when the microeconomic adjustment underlying an aggregate variable is lumpy, conventional VAR procedures often lead the researcher to conclude that there is less persistence than there really is. The extent to which persistence is underestimated decreases with the level of aggregation: linear models capture no persistence when applied to an individual series while the bias vanishes completely when they are applied to a series that aggregates infinitely many agents. Interestingly, convergence is very slow: the bias is likely to be present in general for sectoral data and, quite often, for aggregate series as well. For example, even in the case of the U.S. Consumer Price Index, that aggregates approximately 80,000 prices, the bias turns out to be large, with the estimated half-life of shocks biased downward by approximately 40%.

We propose a procedure to correct for the bias and provide two detailed applications. In the first application, we explain why estimates for the speed of adjustment of sectoral prices obtained using approaches tailored to the underlying lumpy behavior are much lower than those obtained with standard linear time-series models, thereby solving a puzzling finding in Bils and Klenow (2004). We also show that linear time series models deliver estimates in line with those obtained with nonlinear methods once the linear methods are applied correcting for the "missing persistence bias".

Our second application revisits Boivin, Giannoni and Mihov's (2009) finding that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks (see also Mackowiak, Moench and Wiederholt, 2009). In this case we show that once we correct for the missing persistence bias the responses of inflation to both types of shocks look very similar.

The intuition underlying our main result follows from comparing the impulse response of the true nonlinear model that includes lumpy adjustment with the impulse response of a linear approximation, in the simple case of *one* agent and i.i.d. shocks, so that the agent's optimal response every time it acts is to adjust by the sum of shocks that accumulated since the last time it adjusted. We then have that the agent responds in period $t+k$ to a shock that took place in period t only if the agent adjusted in $t+k$ and did not adjust in all periods between t and $t+k-1$. It follows that the average response in $t+k$ to a shock that took place in t is equal to the probability of having to wait exactly k periods until the first opportunity to adjust after the shock takes place. In the simple case where the arrival process that determines when adjustments take place follows a geometric distribution, as in the discrete time version of the Calvo (1983) model, the nonlinear impulse response will be identical to that of an AR(1) process, with persistence parameter equal to the probability of not adjusting in a given period.

Consider next the impulse response obtained using a linear time-series model. This response will depend on the correlations between the agent's actions at different points in time. If the agent did not adjust in one of the periods under consideration, there is no correlation since at least one of the variables entering the correlation is exactly zero. The correlation will also be zero when the agent adjusted at both points in time because the agent's actions reflect shocks in non-overlapping periods and shocks are uncorrelated. This implies that the impulse response obtained via linear methods will be zero at all strictly positive lags, suggesting immediate adjustment to shocks and therefore no persistence, independent of the true degree of persistence. That is, even though the nonlinear IRF recovers the Rotemberg (1987) result, according to which the aggregate of interest follows an AR(1) process with first-order autocorrelation equal to the fraction of units that remain inactive, the linear IRF implies an i.i.d. process which corresponds to the above mentioned AR(1) process when all units adjust in every period and wrongly suggests instantaneous adjustment to shocks.

The bias falls as aggregation rises because the correlations at leads and lags of the adjustments across individual units are non-zero. That is, the common components in the adjustments of different agents at different points in time provides the correlation that allows the econometrician using linear time-series methods to recover the nonlinear impulse response. The more important this common component is —as measured either by the variance of aggregate shocks relative to the variance of idiosyncratic shocks or the frequency with which adjustments take place— the faster the estimate converges to the value of the persistence parameter as the number of agents grows. While idiosyncratic productivity and demand shocks smooth away microeconomic non-convexities and are often used as a justification for approximating aggregate dynamics with linear models, their presence exacerbates the bias. The fact that in practice idiosyncratic uncertainty is many times larger than aggregate uncertainty, suggests that the problem of missing aggregate dynamics is likely to be prevalent in empirical and quantitative macroeconomic research.

Under quite general assumptions, a stationary process can be approximated by a vector autoregression.[2] It is common to infer the speed of adjustment of the process to the innovations from the VAR estimates. When the true process is linear in the innovations, the impulse responses estimated in this way will capture the actual persistence of shocks. By contrast, a central theme underlying the results in this paper is that when the variable of interest aggregates over units with lumpy adjustment, using a VAR will underestimate the true persistence of shocks. The (reduced form) shocks inferred from the VAR estimation differ systematically from the true underlying shocks, and the aggregates of interest respond faster to these estimated shocks than to the true shocks.

The remainder of the paper is organized as follows. Section 2 presents the Rotemberg (1987) equivalence result that justifies using linear time-series methods to estimate the dynamics for ag-

[2] The theoretical underpinning for this statement is Wold's representation result, see Ash and Gardner (1975) for an insightful discussion.

gregates with lumpy microeconomic adjustment, as long as the number of units in the aggregate is infinite. Section 3 begins by presenting the missing persistence bias that arises when the number of units considered is finite. Next we describe approaches to correct for the bias. We also study various extensions of the baseline model, showing that the bias continues being significant. Section 6 studies two detailed applications and Section 7 concludes.

2 Linear Time-Series Models and the Calvo-Rotemberg Limit

Regardless of whether the final goal is to have a reduced form characterization of aggregate dynamics, or whether this is an intermediate step in identifying structural parameters, or whether it is just a metric to assess the performance of a calibrated model, it is common that researchers in macroeconomics at some key stage estimate an equation of the form:

$$a(L)\Delta y_t = \varepsilon_t, \tag{1}$$

where Δy represents the change in the log of some aggregate variable of interest, such as a price index, the level of employment, or the stock of capital; ε is an i.i.d. innovation and $a(L) \equiv 1 - \sum_{k=1}^{p} a_k L^k$, where L is the lag operator and the a_is are fixed parameters.

The question that concerns us here is whether the estimated $a(L)$ captures the true dynamics of the system when the underlying microeconomic variables exhibit lumpy adjustment behavior. We show that unless the effective number of underlying micro units is implausibly large, the answer is 'no'.

We setup the basic environment by constructing a simple model of microeconomic lumpy adjustment. Let y_{it} denote the variable of concern at time t for agent i and y^*_{it} be the level the agent chooses if it adjusts in period t (the 'reset value' of y). We will have that:

$$\Delta y_{it} = \xi_{it}(y^*_{it} - y_{it-1}), \tag{2}$$

where $\xi_{it} = 1$ if the agent adjusts in period t and $\xi_{it} = 0$ if not.

From a modeling perspective, discrete adjustment entails two basic features. First, periods of inaction are followed by abrupt adjustments to accumulated imbalances. Second, the likelihood of an adjustment increases with the size of the imbalance and is therefore state dependent. While the second feature is central for the macroeconomic implications of state-dependent models, it is not needed for the point we wish to raise in this paper. We therefore suppress it in this section and consider it only when analyzing extensions in Section 3. That is, the special model we consider in this section corresponds to that in Calvo (1983) with:

$$\begin{aligned} \Pr\{\xi_{it} = 0\} &= \rho, \\ \Pr\{\xi_{it} = 1\} &= 1 - \rho. \end{aligned} \tag{3}$$

It follows from (3) that the *expected* value of ξ_{it} is $1-\rho$. When ξ_{it} is zero, the agent experiences inaction; when its value is one, the unit adjusts so as to eliminate the accumulated imbalance. We assume that ξ_{it} is independent of $(y^*_{it} - y_{it-1})$ —this is the simplification that Calvo (1983) makes vis-a-vis more realistic state dependent models— and therefore have:

$$\mathrm{E}[\Delta y_{it} \mid y^*_{it}, y_{it-1}] = (1-\rho)(y^*_{it} - y_{it-1}), \tag{4}$$

so that ρ represents the degree of *inertia* of Δy_{it}. When ρ is large, the unit adjusts on average by a small fraction of its current imbalance and the expected half-life of shocks is large. Conversely, when ρ is small, the unit is expected to react promptly to any imbalance.

Let us now consider the behavior of aggregates. Given a set of weights w_i, $i = 1, 2, .., n$, with $w_i > 0$ and $\sum_{i=1}^n w_i = 1$, we define the *effective number of units*, N, as the inverse of the Herfindahl index:

$$N \equiv \frac{1}{\sum_{i=1}^n w_i^2}.$$

When all units contribute the same to the aggregate ($w_i = 1/n$) we have $N = n$, otherwise the effective number of units can be substantially lower than the actual number of units.

We can now write the aggregate at time t, y^N_t, as:

$$y^N_t \equiv \sum_{i-1}^n w_i y_{it}.$$

Similarly we define the value of the aggregate reset value, y^{N*}_t, as

$$y^{N*}_t \equiv \sum_{i-1}^n w_i y^*_{it}.$$

Technical Assumptions (Shocks)

Let $\Delta y^*_{it} \equiv v^A_t + v^I_{it}$, where the absence of a subindex i denotes an element common to all units. We assume:

1. The v^A_t's are i.i.d. with mean μ_A and variance $\sigma^2_A > 0$.

2. The v^I_{it}'s are independent (across units, over time, and with respect to the v^A's), identically distributed with zero mean and variance $\sigma^2_I > 0$.

3. The ξ_{it}'s are independent (across units, over time, and with respect to the v^A's and v^I's), identically distributed Bernoulli random variables with probability of success $\rho \in (0, 1]$. ∎

As Rotemberg (1987) showed, when N goes to infinity, equation (4) for Δy^∞ becomes:

$$\Delta y^\infty_t = (1-\rho)(y^{\infty*}_t - y^\infty_{t-1}). \tag{5}$$

Taking first differences yields

$$\Delta y_t^{\infty} = \rho \Delta y_{t-1}^{\infty} + (1-\rho)\Delta y_t^{\infty *}, \tag{6}$$

which is the analog of Euler equations derived from a simple quadratic adjustment cost model applied to a representative agent.[3]

This is a powerful result which lends substantial support to the standard practice of approximating the aggregates as if they were generated by a simple linear model. What we show below, however, is that while this approximation may be good for some purposes, it can be particularly bad when it comes to motivating VAR estimation of aggregate dynamics.

Before doing so, let us close the loop by recovering equation (1) in this setup. For this, let us momentarily relax the Technical Assumptions 1 and 2, allowing for persistence in the v_t^A and v_{it}^I's, so that the change in the aggregate reset value of y, $\Delta y^{\infty *}$, is generated by:

$$b(L)\Delta y_t^{\infty *} = \varepsilon_t,$$

where the ε_t's are i.i.d. and $b(L) \equiv 1 - \sum_{i=1}^q b_i L^i$ defines a stationary AR(q) for $\Delta y^{\infty *}$. Assuming Technical Assumption 3 holds we have

$$\Delta y_t^{\infty} = \rho \Delta y_{t-1}^{\infty} + (1-\rho)\Delta y_t^{\infty *},$$

which combined with the AR(q) specification for $\Delta y^{\infty *}$ yields

$$(1-\rho L)b(L)\Delta y_t^{\infty} = (1-\rho)\varepsilon_t.$$

Comparing this expression with (1) we conclude that

$$a(L) = b(L)\frac{(1-\rho L)}{1-\rho}.$$

The bias we highlight in this paper comes from a severe downward bias in the (explicit or implicit) estimate of ρ, resulting in an estimate for $a(L)$ that misses significant dynamics. In the next section we simplify the exposition and set $b(L) \equiv 1$, as in the case considered by the Technical Assumptions. We consider the general case in Section 3.5.

3 The Missing Persistence Bias

The effective number of units, N, in any real world aggregate is not infinity. The question that concerns us in this section is whether N is sufficiently large so that the limit result provides a good approximation.

[3]Using quadratic loss functions in economics was initiated by Holt et al. (1961) and continued by Tinsley (1971), Sims (1974) and Sargent (1978). For the proof, see Appendix E.

Our main proposition states that the answer to this question depends on parameter values, in particular, on the relative importance of aggregate and idiosyncratic shocks, the effective number of agents, and the frequency of adjustment. When any of these is small, the bias can remain significant even at the economy-wide level. We argue that this is likely to be the case for various aggregates with lumpy microeconomic adjustment in the U.S. and, by extension, for smaller economies and sectoral data.

3.1 The Theory

We ask whether estimating (6) with an effective number of units equal to N instead of infinity yields a consistent (as T goes to infinity) estimate of ρ, when the true microeconomic model is described by (2) and (3). The following proposition answers this question by providing an explicit expression for the bias as a function of the parameters characterizing adjustment probabilities and shocks (ρ, μ_A, σ_A and σ_I) and N.

Proposition 1 (Aggregate Bias)

Let $\hat{\rho}$ denote the OLS estimator of ρ in

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + e_t. \tag{7}$$

Let T denote the time series length. Then, under the Technical Assumptions, $\text{plim}_{T\to\infty}\hat{\rho}$ *depends on the weights w_i only through N and*

$$\text{plim}_{T\to\infty}\hat{\rho}^N = \frac{K}{1+K}\rho, \tag{8}$$

with

$$K \equiv \frac{\frac{1-\rho}{1+\rho}(N-1) - \left(\frac{\mu_A}{\sigma_A}\right)^2}{1 + \left(\frac{\sigma_I}{\sigma_A}\right)^2 + \frac{1+\rho}{1-\rho}\left(\frac{\mu_A}{\sigma_A}\right)^2}. \tag{9}$$

It follows that:

$$\lim_{N\to\infty} \text{plim}_{T\to\infty}\hat{\rho}^N = \rho. \tag{10}$$

Proof See Appendix C and E. ∎

Equation (10) in the proposition restates Rotemberg's (1987) result. Yet here we are interested in the value of $\hat{\rho}$ before the limit is reached. That is, we would like to assess the value of K.

The bias drops as the effective number of units in the aggregate being considered rises and as the relative importance of aggregate to idiosyncratic shocks rises. Other factors that contribute to slow convergence is a larger drift (in absolute value) in the process driving the reset variable y^*, and a larger degree of inertia as captured by the fraction of agents that do not adjust in any given period, ρ.

3.2 The bias is large in practice

To put the relevance of this non-limit result in perspective, we consider three examples where lumpy microeconomic adjustment has been well established: employment, prices, and investment. Table 1 reports how the half-life and expected response time of shocks varies for these aggregates with the effective number of units, N.[4] We focus on the T=∞ case for two important reasons: the missing persistence bias is conceptually distinct from the well-known AR(1) finite sample bias[5] and in most realistic applications (including our empirical applications in Section 4) the missing persistence bias is an order of magnitude larger than finite sample bias.[6]

Table 1: SLOW CONVERGENCE

Estimated Half-Life of Shocks and Expected Response Time

Aggregate	Frequency	Effective number of agents (N)						
		100	400	1,000	4,000	10,000	40,000	∞
Prices	monthly	0.257	0.464	0.767	1.744	**2.699**	3.886	4.595
Employment	quarterly	0.373	0.663	0.912	**1.197**	1.287	1.338	1.357
Investment	annual	0.179	0.356	**0.582**	1.333	2.167	3.397	4.265
Prices	monthly	0.072	0.290	0.681	2.049	**3.415**	5.121	6.142
Employment	quarterly	0.184	0.541	0.879	**1.275**	1.401	1.474	1.500
Investment	annual	0.021	0.167	**0.436**	1.466	2.653	4.418	5.666

First three rows show the reported half-life. The half-life is inferred from estimation of (7), which is $-\log 2/\log\hat{\rho}_\infty$ with $\hat{\rho}_\infty \equiv \mathrm{plim}_{T\to\infty}\hat{\rho}$ obtained from Proposition 1. The fourth to sixth rows show results when the expected response time (ERT) is the measure of persistence. For an AR(1), the ERT is $\hat{\rho}_\infty/(1-\hat{\rho}_\infty)$ (see Appendix D). Parameters for prices: $\rho = 0.86$, $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$. Parameters for employment: $\rho = 0.60$, $\mu_A = 0.005$, $\sigma_A = 0.03$, $\sigma_I = 0.25$. Parameters for investment: $\rho = 0.85$, $\mu_A = 0.12$, $\sigma_A = 0.056$, $\sigma_I = 0.50$. Numbers in boldface correspond, approximately, to the effective number of units for U.S. aggregates (CPI for prices, non-farm business sector for employment and investment).

The results for prices, reported in the first row in Table 1, assume $\rho = 0.86$, in line with the median frequency of price adjustments for regular prices reported in Klenow and Kryvtsov (2008).[7] Values for μ_A and σ_A are taken from Bils and Klenow (2004), while σ_I is consistent with the value

[4] See Appendix D for the definition and main properties of the expected response time.

[5] See Hamilton 1994 pp 216 for a textbook treatment.

[6] Monte-Carlo results confirming this statement are available upon request.

[7] The average over the eight median frequencies reported by Nakamura and Steinsson (2008) for regular price changes suggest taking $\rho = 0.89$ which leads to a larger bias.

estimated in Caballero et al (1997).[8] The table shows that the bias remains significant even for $N = 10,000$, which corresponds, approximately, to the effective number of prices used to calculate the CPI.[9] In this case, the main reason for the bias is the high value of σ_I/σ_A.

The second row in Table 1 reports the results for aggregate U.S. employment. We use the parameters estimated by Caballero, Engel, and Haltiwanger (1997) with quarterly Longitudinal Research Datafile (LRD) data for μ_A, σ_A, σ_I and ρ. The second row in Table 1 suggests that with $N = 3,683$, which is the effective size of employment in the non-farm business sector in 2001, the bias is only slightly above 10%. However, note that when $N = 100$, which corresponds to the average effective number of establishments in a typical two-digit sector of the LRD, the estimate half-life of shocks is less than one third of the actual half-life.

Finally, the third row in Table 1 reports the estimates for equipment investment, the most sluggish of the three series. The estimate of ρ, μ_A and σ_A, are from Caballero, Engel, and Haltiwanger (1995), and σ_I is consistent with that found in Caballero et al. (1997).[10] Here the bias remains very large and significant throughout. In particular, when $N = 986$, which corresponds to the effective number of establishments for capital weights in the U.S. Non-Farm Business sector in 2001, the estimated half-life of a shock is only 14% of the true half-life or, equivalently, the estimated frequency of adjustment, $1-\rho$, is more than four times the true frequency. The reasons for this is the combination of a high ρ, a high μ_A (mostly due to depreciation) and a large σ_I (relative to σ_A).

Summing up, the missing persistence bias is large at the sectoral level for inflation, employment and investment. Furthermore, linear time-series models will miss a substantial part of the dynamic behavior of U.S. inflation and investment at the aggregate level as well. The true half-life of a shock is close to twice its estimate for inflation and more than seven times its estimate for investment. Even though the setting we have used to gauge the magnitude of the bias is quite simple, in Section 3.5 we show that these conclusions extend to more general settings.

3.3 What is behind the bias and slow convergence?

Having established the proposition and the practical relevance of the bias, let us turn to the intuition behind the proof of the proposition. We do this in two steps. We first describe the genesis of the bias,

[8] To go from the σ_I computed for employment in Caballero et al. (1997) to that of prices, we note that if the demand faced by a monopolistic competitive firm is isoelastic, its production function is Cobb-Douglas, and its capital fixed (which is nearly correct at high frequency), then (up to a constant):

$$p^*_{it} = (w_t - a_{it}) + (1-\alpha_L)l^*_{it}$$

where p^* and l^* denote the logarithms of frictionless price and employment, w_t and a_{it} are the logarithm of the nominal wage and productivity, and α_L is the labor share. It is straightforward to see that as long as the main source of idiosyncratic variance is demand, which we assume, $\sigma_{I_{p^*}} \simeq (1-\alpha_L)\sigma_{I_{l^*}}$.

[9] The median (mean) total number of observations per month between 1988:02 and 2007:12 is 66,582 (67,428). The median (mean) *effective* number of observations per month during this period is 10,328 (10,730).

[10] To go from the σ_I computed for employment in Caballero et al. (1997) to that of capital, we note that if the demand faced by a monopolistic competitive firm is isoelastic and its production function is Cobb-Douglas, then $\sigma_{I_{k^*}} \simeq \sigma_{I_{l^*}}$.

which can be seen most clearly when $N = 1$. We then show why, for realistic parameter values, the extreme bias identified for $N = 1$ vanishes very slowly as N grows.

3.3.1 The genesis of the bias

Let us set $\mu_A = 0$. From (8) we have that when $N = 1$, regardless of the true value of ρ,

$$\text{plim}_{T\to\infty}\hat{\rho} = 0. \tag{11}$$

That is, a researcher that uses a linear model to infer the speed of adjustment from the series for one unit will conclude that adjustment is infinitely fast independent of the true value of ρ. Of course, few would estimate a simple AR(1) for a series of one agent with lumpy adjustment, but the point here is not to discuss optimal estimation strategies for lumpy models but to illustrate the source of the bias step-by-step. The case $N = 1$ is a convenient first step in this process.

The key point to notice is that when adjustment is lumpy, the correlation between this period's and the previous period's adjustment is zero, independently of the true value of ρ. To see why this is so, consider the covariance of Δy_t and Δy_{t-1}, noting that, because adjustment is complete whenever it occurs, we may re-write (2) as:

$$\Delta y_t = \xi_t \sum_{k=0}^{l_t-1} \Delta y^*_{t-k} = \begin{cases} \sum_{k=0}^{l_t-1} \Delta y^*_{t-k} & \text{if } \xi_t = 1, \\ 0 & \text{if } \xi_t = 0, \end{cases} \tag{12}$$

where l_t denotes the number of periods, as of period t, since the last adjustment took place. So that $l_t = 1$ if the unit adjusted in period $t-1$, 2 if it did not adjust in $t-1$ and adjusted in $t-2$, and so on.

Table 2: CONSTRUCTING THE MAIN COVARIANCE

Adjust in $t-1$	Adjust in t	Δy_{t-1}	Δy_t	Contribution to $\text{Cov}(\Delta y_t, \Delta y_{t-1})$
No	No	0	0	$\Delta y_t \Delta y_{t-1} = 0$
No	Yes	0	Δy^*_t	$\Delta y_t \Delta y_{t-1} = 0$
Yes	No	$\sum_{k=0}^{l_{t-1}} \Delta y^*_{t-1-k}$	0	$\Delta y_t \Delta y_{t-1} = 0$
Yes	Yes	$\sum_{k=0}^{l_{t-1}} \Delta y^*_{t-1-k}$	Δy^*_t	$\text{Cov}(\Delta y_{t-1}, \Delta y_t) = 0$

There are four scenarios to consider when constructing the key covariance (see Table 2). If there is no adjustment in this and/or the last period (three scenarios), then the product of this and last period's adjustment is zero, since at least one of the adjustments is zero. This leaves the case of adjustments in both periods as the only possible source of non-zero correlation between consecutive adjustments. Conditional on having adjusted both in t and $t-1$, we have

$$\text{Cov}(\Delta y_t, \Delta y_{t-1} \mid \xi_t = \xi_{t-1} = 1) = \text{Cov}(\Delta y^*_t\,,\, \Delta y^*_{t-1} + \Delta y^*_{t-2} + \cdots + \Delta y^*_{t-l_{t-1}-1}) = 0,$$

since adjustments in this and the previous period involve shocks occurring during non-overlapping time intervals. Every time the unit adjusts, it catches up with all previous shocks it had not adjusted to and starts accumulating shocks anew. Thus, adjustments at different moments in time are uncorrelated.

The case $N = 1$ is also useful to compare the impulse responses inferred from linear models with those obtained from first principles. We define the latter via:

$$I_k \equiv \mathrm{E}_t \left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*} \right].$$

It follows from Proposition 1 that the impulse response of Δy to Δy^* inferred from a linear time-series model estimated for an individual series of Δy will be equal to one upon impact and zero for higher lags.

To calculate the correct impulse response, we note that Δy_{t+k} responds to Δy_t^* if and only if the first time the unit adjusted after the period t shock took place is in period $t+k$. It also follows from our Technical Assumptions that in this event the response is one-for-one. Thus

$$I_k = \Pr\{\xi_t = 0, \xi_{t+1} = 0, ..., \xi_{t+k-1} = 0, \xi_{t+k} = 1\} = (1-\rho)\rho^k.$$

This is the IRF for an AR(1) process obtained for *aggregate* inflation in the standard Calvo model (see, for example, Section 3.2 in Woodford, 2003).[11]

What happened to Wold's representation, according to which any process that is stationary and non-deterministic admits an (eventually infinite) MA representation? Why is Wold's representation in this case an i.i.d. process, suggesting an infinitely fast response to shocks, independent of the true persistence of shocks?

In general, Wold's representation is a distributed lag of the one-step-ahead *linear* forecast errors for the process. In the case we consider here we have $\mathrm{E}[\Delta y_t \Delta y_{t+1}] = 0$ and therefore $\Delta y_{t+1} - \mathrm{E}[\Delta y_{t+1} | \Delta y_t] = \Delta y_{t+1}$ so that the Wold innovation at time $t+1$, Δy_{t+1}, differs from the innovation of economic interest, Δy_{t+1}^*.

Wold's representation does not necessarily capture the entire process but only its first two moments. If higher moments are relevant, as is generally the case when working with variables that involve lumpy adjustment, the response of the process to the innovation process in Wold's representation will not capture the response to the economic innovation of interest.

3.3.2 Slow convergence

We have characterized the two extremes. When $N = 1$, the bias is maximum; when $N = \infty$ there is no bias. Next we explain how aggregation reduces the bias, and then study the speed at which

[11] As discussed in Caballero and Engel (2007), the impulse response for an individual unit and the corresponding aggregate will be the same for a broad class of macroeconomic models, including the one specified by the Technical Assumptions in Section 2.

convergence occurs.

For this purpose, we begin by writing $\hat{\rho}$ as an expression that involves sums and quotients of four different terms:

$$\mathrm{plim}_{T\to\infty}\hat{\rho} = \frac{\mathrm{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)}{\mathrm{Var}(\Delta y_t^N)} = \frac{\sum_i w_i^2 \mathrm{Cov}(\Delta y_{1,t}, \Delta y_{1,t-1}) \;+\; \sum_{i\neq j} w_i w_j \mathrm{Cov}(\Delta y_{1,t}, \Delta y_{2,t-1})}{\sum_i w_i^2 \mathrm{Var}(\Delta y_{1,t}) \;+\; \sum_{i\neq j} w_i w_j \mathrm{Cov}(\Delta y_{1,t}, \Delta y_{2,t})},$$

and since $N = 1/\sum_i w_i^2$ and $\sum_i w_i = 1$:

$$\mathrm{plim}_{T\to\infty}\hat{\rho} = \frac{N\mathrm{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) \;+\; N(N-1)\mathrm{Cov}(\Delta y_{it}, \Delta y_{j,t-1})}{N\mathrm{Var}(\Delta y_{it}) \;+\; N(N-1)\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt})}, \tag{13}$$

where the subindices i and j in Δy denote two different units. Table 3 provides the expressions for the four terms that enter in the calculation of $\hat{\rho}$.

Table 3: CONSTRUCTING THE FIRST ORDER CORRELATION

	$\mathrm{Cov}(\Delta y_{it}, \Delta y_{i,t-1})$	$\mathrm{Cov}(\Delta y_{it}, \Delta y_{j,t-1})$	$\mathrm{Var}(\Delta y_{it})$	$\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt})$
Lumpy ($\mu_A = 0$):	0	$\frac{1-\rho}{1+\rho}\rho\sigma_A^2$	$\sigma_A^2 + \sigma_I^2$	$\frac{1-\rho}{1+\rho}\sigma_A^2$
Lumpy ($\mu_A \neq 0$):	$-\rho\mu_A^2$	$\frac{1-\rho}{1+\rho}\rho\sigma_A^2$	$\sigma_A^2 + \sigma_I^2 + \frac{2\rho}{1-\rho}\mu_A^2$	$\frac{1-\rho}{1+\rho}\sigma_A^2$

If $N = 1$, only the two within-agent terms remain, one in the numerator and one in the denominator. Since the covariance in the numerator is zero,[12] $\hat{\rho}$ is zero as well. This drag on $\hat{\rho}$ remains present as N grows, but its relative importance declines since the between-agents covariances in the numerator and denominator are multiplied by terms of order N^2. This means that the reduction of the bias must come from the between-agents correlations at leads and lags, captured by the second expression in the numerator and denominator. The expression in the numerator is positive because not all individual units react to common shocks at the same time. The expression in the denominator is positive, because some do react at the same time. Either way, it is clear that these expressions are proportional to the variance in aggregate shocks only. In fact, as summarized in the first row of Table 3:

$$\mathrm{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) = \frac{1-\rho}{1+\rho}\rho\sigma_A^2,$$

$$\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt}) = \frac{1-\rho}{1+\rho}\sigma_A^2,$$

and we see that the ratio of the two between-agents covariance terms is indeed ρ. When N goes to infinity, it is this ratio that dominates $\hat{\rho}$.

While these between-agents terms are proportional to the variance of aggregate shocks only,

[12] For simplicity we continue assuming $\mu_A = 0$.

the within-agent responsible for the biases are proportional to total uncertainty. In particular, the denominator of (13) is

$$\mathrm{Var}(\Delta y_{1,t}) = \sigma_A^2 + \sigma_I^2,$$

which cannot be compensated by the within-agent covariance in the numerator since this is equal to zero for the reasons described earlier. Thus $\hat{\rho}$ remains small even for large values of N.

Aside from the relative importance of idiosyncratic shocks for the bias, we see from the expression for K in Proposition 1 that the bias is larger when the drift is different from zero and when persistence is high. The latter is intuitive: When ρ is high, the between-agents covariances are small since adjustments across units are further apart, thus a larger number of units are required for these terms to dominate in the calculation of $\hat{\rho}$.

To understand the impact of the drift on convergence, we must explain why the covariance between Δy_t and Δy_{t-1} for a given unit is negative when $\mu_A \neq 0$ and why the variance term increases with $|\mu_A|$ (see the second row in Table 3). To provide the intuition for the negative covariance, assume $\mu_A > 0$ (the argument is analogous when $\mu_A < 0$) and note that the unconditional expectation of Δy_t is equal to μ_A, which corresponds to expected adjustment when adjusting in consecutive periods (the intuition is straightforward, see Appendix C for a formal proof). Expected adjustment when adjusting after more than one period are larger than μ_A. It follows that a value of Δy_t above average suggests that it is likely that the agent did not adjust in $t-1$, implying that Δy_{t-1} is likely to be smaller than average. Similarly, a value of Δy_t below average suggests that it is likely that the agent adjusted in period $t-1$, and Δy_{t-1} is likely to be larger than average in this case.

The reason why the variance term increases when $\mu_A \neq 0$ is that the dispersion of accumulated shocks is larger in this case, because by contrast with the case where $\mu_A = 0$, conditional on adjusting, the average adjustment increases with the number of periods since the unit last adjusted (it is equal to μ_A times the number of periods).

Summing up, linear time-series models use a combination of self- and cross-covariance terms involving units' adjustments to estimate the microeconomic speed of adjustment. Inaction biases the self-covariance terms toward infinitely fast adjustment (and beyond when $\mu_A \neq 0$). It follows that the ability to recover the true value on ρ depends on the cross-covariance terms playing a dominant role. Yet these terms recover ρ thanks to the common components in the adjustment of different units in consecutive periods, thus their contribution when estimating ρ will be smaller when adjustment is less frequent (larger ρ), and when idiosyncratic uncertainty is large relative to aggregate uncertainty.

3.4 Bias Correction

This section studies an approach to correct for the missing persistence bias, based on using a proxy for the reset value y^*. Two alternative approaches—one based on an ARMA representation of Δy_t^N, the other on instrumental variables—are discussed in Appendix A.

So far we have assumed that the sluggishness parameter ρ is estimated using only information on the economic series of interest, y. Yet often the econometrician can resort to a proxy for the reset value y^*. Instead of (7), the estimating equation, which is valid for $N=\infty$, becomes:

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + (1-\rho)\Delta y_t^{*N} + e_t, \tag{14}$$

with some proxy available for the regressor Δy^*.

Equation (14) hints at a procedure for correcting the bias since it tells us what the correct control function to use is to get an unbiased estimate of ρ: use a proxy for innovation in the shock, Δy_t^*. Since the regressors are orthogonal, from Proposition 1 we have that the coefficient on Δy_{t-1} will be biased downward. By contrast, the true speed of adjustment can be estimated directly from the parameter estimate associated with Δy_t^*, as long as the constraint that the sum of the coefficients on both regressors add up to one is not imposed. Of course, the estimate of ρ will be biased if the econometrician imposes the latter constraint. We summarize these results in the following proposition.

Proposition 2 (Bias with Regressors)

With the same notation and assumptions as in Proposition 1, consider the following equation:

$$\Delta y_t^N = \text{const.} + b_0 \Delta y_{t-1}^N + b_1 \Delta y_t^{*N} + e_t, \tag{15}$$

where Δy_t^{*N} denotes the average shock in period t, $\sum w_i \Delta y_{it}^*$. Then, if (15) is estimated via OLS, and K defined in (9),

(i) without any restrictions on b_0 and b_1:

$$\begin{aligned} plim_{T\to\infty}\hat{b}_0 &= \frac{K}{1+K}\rho, && (16)\\ plim_{T\to\infty}\hat{b}_1 &= 1-\rho; && (17)\end{aligned}$$

(ii) imposing $b_0 = 1 - b_1$:

$$plim_{T\to\infty}\hat{b}_0 = \rho - \frac{(1-\rho)^2}{K+1-\rho}.$$

Proof See Appendix C. ∎

Proposition 2 entails the general message that constructing a proxy for the reset variable y^* can be very useful when estimating the dynamics of a macroeconomic variable with lumpy microeconomic adjustment. Also, it is important to avoid imposing constraints that hold only when $N=\infty$. We apply this approach in Section 4.

3.5 Extensions

The Technical Assumptions we made so far in this section allowed for closed form expressions and simple intuitions for the missing persistence bias. In Appendix B we consider the following departures from the assumptions we have made so far: the probability of adjusting is state-dependent, y^* does not follow a random walk, agents' decisions are strategic complements and agents' adjustment decisions are lumpy but spread out over time ('time-to-build'). We show that the missing persistence bias continues be present (and significant) in all of these cases.

4 Applications

In Section 2 we established the existence of the missing persistence bias theoretically, in Section 3 we argued, via simple calibration exercises, that it is likely to be large in practice. In this section we go one step further and present two applications where recent findings on inflation dynamics are overturned once the missing persistence bias is considered.

The pricing literature is a natural context in which to study the relevance of the missing persistence bias because numerous studies over the last decade have shown that at the item level prices adjust infrequently.[13] Both applications provide evidence of the presence of the bias and correct for it using the approach outlined in Section 3.4. To correct for the bias we construct an estimate for the aggregate and sectoral shocks facing retail price-setters, based on establishment level prices. These series are of interest in their own right and can be of use in other applications.

Our first example shows that accounting for the missing persistence bias overturns Bils and Klenow's rejection of the Calvo model from their now classic 2004 paper.[14] We start with this simple example because the assumptions are identical to those underlying the results in Section 3 and because we are able to calculate the exact magnitude of the bias in this case based on the CPI micro database. We show that the bias is substantial and that the bias correction procedure eliminates the bias almost entirely.

In our second application, we turn to recent empirical work using sectoral price data to argue that firms respond faster to sectoral shocks than to aggregate shocks (Boivin, Giannoni and Mihov, 2009; Mackoviak, Moench and Wiederholt, 2009). These results have been interpreted as evidence in favor of rational inattention or imperfect information models of price setting, because they suggest that firms respond more to bigger, more salient shocks. However, we show that once the missing persistence bias is accounted for, there is little evidence that sectoral prices respond faster to sectoral shocks than to aggregate shocks.

[13] For evidence based on the micro database used to calculate the CPI see Bils and Klenow (2004), Nakamura and Steinsson (2008) and Klenow and Kryvtsov (2008).

[14] The findings that follow do not affect the main contribution of their paper, which is to provide broad based evidence on the extent to which U.S. prices are sticky at the microeconomic level.

4.1 Example 1: Solving a Puzzle in Bils-Klenow

Figure 2 in Bils and Klenow's influential 2004 paper (BK in what follows) presents a scatter plot of the frequency of price adjustments, λ_s, estimated from retail level pricing data, and the coefficient ρ_s estimated via OLS from the following regression using the sectoral inflation series π_{st}:

$$\pi_{st} = \rho_s \pi_{s,t-1} + e_{st}. \tag{18}$$

Under the assumptions of the Calvo pricing model considered in Section 3, which are the assumptions considered by BK, we should have that $\hat{\rho}_s$ is approximately equal to $1 - \hat{\lambda}_s$. In contrast, BK find that in all sectors $\hat{\rho}_s$ is smaller than $1 - \hat{\lambda}_s$, with a substantial difference in most cases.

In other words, Figure 2 in BK shows that the persistence of shocks inferred from a linear time-series model estimated with sectoral data is considerably smaller than the true persistence parameter inferred from microeconomic retail pricing data. BK interpret this finding as evidence against the Calvo model. However we show below that the missing persistence bias leads to downward biased estimates of the sectoral ρ_s and that once we correct for this bias the systematic difference between $\hat{\rho}_s$ and $1 - \hat{\lambda}_s$ disappears.[15]

We proceed in three steps. First we calibrate a multisector Calvo model and show that figures obtained from simulating this model look similar to Figure 2 in BK. Next we use the CPI micro database and the reset price inflation methodology of Bils, Klenow and Malin (2012) to estimate sectoral shocks series. We then use the bias correction approach from Section 3.4 to obtain estimates for ρ_s that are immune to the missing persistence bias. We find that the bias correction method does a good job, that is, we find that $\hat{\rho}_s \simeq 1 - \hat{\lambda}_s$.

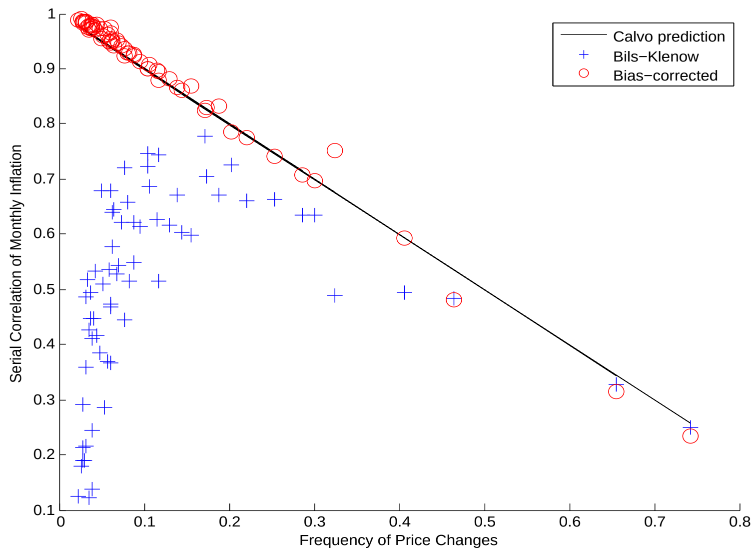
To gauge whether the bias could be an explanation for the BK finding, we first obtain a back of the envelope estimate of whether the magnitude of the bias is quantitatively similar to the magnitude suggested by Figure 2 in BK. Towards this end, we calibrate a multi-sector version of the Calvo model and compare the true adjustment frequencies with those estimated by linear time-series methods using simulated data. We work with the two-digit or "Expenditure class" level of aggregation rather than the ELI level of aggregation used in BK because we will need to estimate underlying shocks when correcting for the bias and this level of aggregation provides a good balance between having a sufficiently large number of sectors and being able to obtain good estimates for underlying shocks.[16] The number of sectors we consider is 66.

The calibration we use is standard and the details are relegated to Appendix H.1. Of course, an important element in our calibration is that we set the number of effective price-setters in each sector to the number observed in the CPI micro database. Our multi-sector model provides a simple laboratory to test whether the missing persistence bias is relevant in this case. The implications

[15] For an alternative explanation for the bias see Le Bihan and Matheron (2012)

[16] We only use representative monthly pricing data in constructing our price indices to be able to measure monthly shocks, which cuts down our underlying sample sizes significantly when compared to using bimonthly data as well. Also, we only chose those sectors for which we could have data for the entire sample period.

Figure 1: Simulation Counterpart to Bils-Klenow figure from Multi-sector Calvo Model



from our simulations are summarized in Figure 1, the prediction of the Calvo model is shown by the solid black line. The BK prediction is shown by the blue crosses. Consistent with BK's results, we find that the estimated persistence of sectoral inflation rates is much lower than is implied by the Calvo model. That is, the blue crosses always lie below the black line (the Calvo prediction) just as Bils and Klenow found using the CPI micro database.

We use the reset price methodology of Bils, Klenow and Malin (2012), applied to each sector separately, to construct a sectoral-specific estimate for the shock (see Appendix G for details).

Following Section 3.4 we implement our bias correction procedure by including our measure of the sectoral shock as an additional control in equation (15):

$$\pi_{st} = \beta_s \pi_{s,t-1} + \gamma_s v_{st} + e_{st}. \tag{19}$$

Proposition 2 implies that if we estimate β and γ in the above equation without imposing any constraints across them, $\widehat{\gamma}_s$ will be an unbiased estimate of the actual fraction of adjusters λ_s.

We first return to our simulated multi-sector Calvo model, estimate the v-shocks using our repeat-price-change methodology and then estimate the above regression sector by sector. The results for each sector are represented by circles in Figure 1, where each circle represents the average of 500 corrected estimates for $1-\lambda_s$ obtained via simulations, based on estimating a linear time series model for sectoral inflation data. All estimates now lie close to the Calvo prediction (the solid line).

Next we implement the bias correction approach using micro data on prices from the BLS. We use the CPI research database which contains individual price observations for the thousands of non-shelter items underlying the CPI over the sample period 1988:03-2007:12. Prices are collected monthly for all items only in New York, Los Angeles and Chicago, and we restrict our analysis to these cities to ensure the representativeness of our sample.[17] The database contains thousands of individual "quote-lines" with price observations for many months. In our data set, an average month contains approximately 12,000-15,000 different quote-lines. Quote-lines are the highest level of disaggregation possible and correspond to an individual item at a particular outlet. An example of a quote-line collected in the research database is a16 oz bag of frozen corn at a particular Chicago outlet.

Much of the recent literature has discussed the difference between sales, regular price changes and product substitutions. We exclude sales following Eichenbaum, Jaimovich, and Rebelo (2012) and Kehoe and Midrigan (2012), who argue that the behavior of sales is often significantly different from that of regular or reference prices and that regular prices are likely to be the important object of interest for aggregate dynamics. We exclude product substitutions because these require a judgement on what portion of a price change is due to quality adjustment and which component is a pure price change. This introduces measurement error in the calculation of price changes at the time of product substitution. Bils (2009) shows that these errors can be substantial.[18]

As a first step we replicate Bils and Klenow's (2004) results for our 66 sectors. First we estimate equation (18) using the micro data, and denote the implied frequency of adjustment estimates as $\lambda_s^{\text{VAR}} = 1 - \widehat{\beta}_s$. As in Bils and Klenow (2004), we find that $\widehat{\beta}_s \ll 1 - \lambda_s^{\text{micro}}$, where λ_s^{micro} denotes the true frequency of adjustment, estimated from the micro level quote-lines. Next we estimate equation (19) using our constructed shock measure, v_{st}, based on the repeat price-change approach outlined above.

We denote the coefficient on our sectoral shock measure by λ_s^c, where the superindex c stands for "corrected". To gauge the extent to which the λ_s^c corrects the missing persistence bias, we regress the change in estimated speed of adjustment we achieve in a given sector on the magnitude of the bias (which in this particular case is known). That is, we estimate by OLS the following equation:

$$(\lambda_s^c - \lambda_s^{\text{VAR}}) = \alpha + \eta(\lambda_s^{\text{micro}} - \lambda_s^{\text{VAR}}) + \text{error}.$$

Here η is the coefficient of interest as it captures the extent to which our bias correction actually decreases the bias. If the bias reduction is large but unrelated to the magnitude of the bias, the estimated value of α will be large while η won't be significantly different from zero. By contrast, if the bias reduction is proportional to the actual bias, we expect an estimate of η that is significantly

[17] The most representative sample would be to use all bimonthly observations, but then many price changes are potentially missing. Some items are sampled monthly outside of New York, Los Angeles and Chicago, but these items are not representative, so we restrict our monthly analysis to these three cities.

[18] Nevertheless, we have also repeated the analysis including product substitutions and found similar results.

positive, taking values close to one if the bias completely disappears.

Table 4: Bias-Correction Estimation

	Multi-sector Calvo Model (simulations)	CPI database (actual data)
η	1.038***	1.000***
	(0.012)	(0.028)
Constant	-0.003	-0.063***
	(0.005)	(0.024)
Observations	66	66
R-squared	0.99	0.95

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p< 0.1

Table 4 shows the estimates we obtain. Both in the multi-sector Calvo simulation and with the CPI database, our bias correction strategy comes very close to eliminating the bias entirely. For the CPI data, the estimated value of η is not statistically different from one. This suggests that the departure from the Calvo model found in Figure 2 in BK is likely driven by the missing persistence bias. Thus, this example shows that the bias is relevant at the sectoral level and that through the use of microeconomic data our control function approach can be used to overcome this bias.

4.2 Example 2: Faster response to sectoral shocks than to aggregate shocks?

The theoretical literature on sticky-information and costly observation models points out that there is no reason why prices should adjust equally fast to different types of shocks. For example, Boivin, Gianonni and Mihov (2009) (henceforth BGM) provide empirical evidence that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks, which is consistent with both of these classes of models. However, differential speed of adjustment to shocks at different levels of aggregation could also signal the presence of the missing persistence bias. We explore this possibility next and show that the difference in speed of adjustment disappears once we correct for the bias.

To understand BGM's approach, we must first introduce some terminology. Define Π_t as a column vector with monthly sectoral inflation rates in period t, for sectors 1 through S, based on data from the BEA and the PPI, where S denotes the number of sectors. BGM assume that Π_t can be decomposed into the sum of small number K of common factors, C_t, and a sectoral component, e_t:

$$\Pi_t = \Lambda C_t + e_t, \tag{20}$$

where Λ denotes an SxK matrix of factor loadings that are allowed to differ across sectors, while C_t and e_t are Kx1 and Sx1 matrices. This formulation allows them to disentangle the fluctuations in sectoral inflation rates due to the macroeconomic factors—represented by the common components C_t which sector specific weights—from those due to sector-specific conditions represented by the term e_t.

BGM extract K principal components from the large data set Π_t to obtain consistent estimates of the common factors.[19] Next they regress each sectoral inflation series on the common factors,[20] denoting the predicted aggregate component, $\lambda_i' C_t$, by π_{st}^{agg}, and the residual that captures the sector-specific component, e_{st}, by π_{st}^{sect}:

$$\pi_{st} = \lambda_s' C_t + e_{st} = \pi_{st}^{\text{agg}} + \pi_{st}^{\text{sect}}. \tag{21}$$

To calculate IRFs with respect to the common and sectoral shocks, BGM fit separate AR(13) processes to π_{st}^{agg} and π_{st}^{sect} series and measure the persistence of shocks by the sum of the 13 AR coefficients. For example, for the processes considered under the Technical Assumptions this sum is equal to ρ.

Table 5: BGM'S RESULTS WITH OUR CPI SERIES

Sum of AR coefficients for AR(13)

	π_{st}^{agg}	π_{st}^{sect}
Average over 66 series	0.45	−0.11
Median over 66 series	0.64	−0.04

To start, we reproduce their benchmark results using our 66 series and compare the results we obtain to what BGM found using a different time period and data.[21] Table 5 shows we find similar results to BGM when we replicate their methodology in the CPI data.[22] We report the mean and median of the same persistence measure used by BGM.

Even though the persistence measures we obtain for the response to aggregate shocks are somewhat smaller than those reported by BGM, the difference between the persistence of the aggregate and sectoral components of sectoral inflation are similar to those in BGM. There is clear evidence of significant persistence for the former and no evidence of persistence for the latter. A similar conclusion was reached by Mackoviak, Moench and Wiederholt (2011) using CPI data and a different

[19] Stock and Watson (2002) show that the principal components consistently recover the space spanned by the factors when S is large and the number of principal components used is at least as large as the true number of factors.

[20] BGM allow C_t to follow an AR process. We allow for this possibility and allow C_t to have 6 lags in our baseline estimation. We have also tried different specifications where we allow for either 0 or 12 lags of C_t and found similar results.

[21] There are a number of differences between our sample and BGM's. The most notable difference is that BGM use disaggregated information on both prices and quantities whereas we just use information on prices. The other main differences are over sample period (BGM use the sample period 1976-2005 whereas we use data over the time period 1988-2007) and in the number of series used (BGM use 600 in their baseline whereas we use 66).

[22] We report results that assume there are 4 common factors, with three lags in each of these factors. Results are robust for reasonable deviations from these assumptions.

methodology. Both BGM and Mackoviak et al (2009) conclude that this difference in persistence is strong evidence in favor of sticky-information models. We revisit this conclusion next.[23]

We begin by noting that BGM's persistence measure is calculated by first regressing each component on lags of itself. Since the underlying prices adjust infrequently and there are not many prices underlying these sectoral inflation series, BGM's results could be driven by the missing persistence bias.

To investigate this hypothesis, we use the same shock measures that we computed from CPI micro data that were discussed in depth in Section 4.1. That is, we have data for 66 sectoral inflation series from the CPI for the period 1988:03-2007:12, together with the corresponding series of innovations (the v's from Section 4.1).

Define V_t as the Sx1 vector with the period t sectoral shock measures. Our proxy for the common components of the aggregate shock are the first K principal components of V, denoted by m_t^k, $k = 1, 2, ..., K$. To decompose the v_{st} into the sum of an aggregate and a sectoral component we regress these shocks on the common factors and their lags and denote the residual by x_{st}:[24]

$$v_{st} = \sum_{k=1}^{K} \sum_{j=0}^{J} \gamma_{sj}^{k} m_{t-j}^{k} + x_{st}. \tag{22}$$

The term with double sums on the r.h.s. is the component driven by aggregate shocks, while the residual x_{st} is the component driven by sectoral shocks.

So far we have K aggregate shock components, m_t^k, and a sectoral shock, x_{st}, for each of the 66 sectoral innovation series we obtained from the CPI using reset price inflation in Section 4.1. Next we decompose the sectoral inflation series into two components, one driven by aggregate shocks, the other by sectoral shocks. To do this, we estimate:

$$\pi_{st} = \sum_{k=1}^{K} \eta_s^k(L) m_t^k + \nu_s(L) x_{s,t}, \tag{23}$$

where $\eta_s^k(L) = \sum_{j \geq 0} \eta_{sj} L^j$ and $\nu_s(L) = \sum_{j \geq 0} \nu_{sj} L^j$ denote lag polynomials. We model each $\eta_s^k(L)$ and $\nu_s(L)$ as quotients of two second degree polynomials. The results we obtain are robust to reasonable variations in the order of these polynomials.[25]

The approach we use to correct for the missing persistence bias is based on information that is not included in the sectoral inflation series and therefore we must use a persistence measure that is different from the one used by BGM. We consider the expected response time (see Section 3.2 and

[23] Carlsson and Skans (2012) use firm level information on prices and marginal costs from Sweden, and find that cost pass-through to idiosyncratic cost shocks is much less than one. They interpret this finding as contradicting the predictions of the Rational Inattention Model of Mackowiak and Wiederholt (2009).

[24] Our results are robust to ignoring distributed lags of common components yet we believe it is more realistic to include these components as aggregate shocks might affect sectoral shocks with a lag.

[25] These robustness results are available upon request. We implemented this estimation using the polyest command in Matlab. See http://jp.mathworks.com/help/ident/ref/polyest.html for details.

Appendix D) to each of the K aggregate shocks and summarize the K response times to aggregate shocks by their median:

$$\begin{aligned}
\tau_s^{\text{sec}} &\equiv \sum_{j\geq 0} j\nu_{sj}^k / \sum_{j\geq 0} \nu_{sj}^k, \\
\tau_s^{\text{agg},k} &\equiv \sum_{j\geq 0} j\eta_{sj}^k / \sum_{j\geq 0} \eta_{sj}^k, \\
\tau_s^{\text{agg}} &\equiv \text{median}_k \tau_{s,k}.
\end{aligned}$$

Because we have a direct proxy for both shocks, our measures of persistence to these shocks are not susceptible to the missing persistence bias.

Table 6: THE RESPONSE OF SECTORAL INFLATION RATES TO AGGREGATE AND IDIOSYNCRATIC SHOCKS

Median of estimated expected response times to shocks

PCs	nlags	agg	sec
		(1)	(2)
2	0	3.63	3.03
		(0.84)	(0.56)
2	3	2.57	2.71
		(0.77)	(0.55)
2	6	3.05	1.77
		(0.86)	(0.51)
2	12	2.79	2.86
		(0.91)	(0.56)
4	0	2.72	2.56
		(0.44)	(0.53)
4	3	1.98	2.53
		(0.44)	(0.54)
4	6	2.12	1.99
		(0.34)	(0.50)
4	12	1.72	2.17
		(0.45)	(0.54)
6	0	1.87	2.51
		(0.38)	(0.50)
6	3	2.00	2.83
		(0.46)	(0.64)
6	6	1.97	2.56
		(0.33)	(0.55)
6	12	2.14	2.24
		(0.33)	(0.56)

The results are shown in Table 6. The numbers we report are medians across sectors. The interquartile ranges (divided by the square root of the number of sectors) are shown in parentheses. We consider 12 possible combinations for the number of principal components (PC) and number of lags (nlags) used on the r.h.s. of (22).

Columns (1) and (2) show that after correcting for the missing persistence bias using the procedure outlined above, there is no significant difference between the estimated response times of sectoral inflation series to aggregate and sectoral shocks. The average difference between corrected estimates is both economically and statistically small (2.38 vs. 2.48 months) and, if anything, the sectoral component of sectoral inflation is more persistent than the aggregate component. We conclude that once one corrects for the missing persistence bias, there is no longer evidence that firms respond differently to aggregate and sectoral shocks.

5 Conclusion

While many (if not most) microeconomic actions are infrequent and lumpy, large idiosyncratic shocks map these discrete microeconomic series into smooth aggregated counterparts. The presumption (either explicit or implicit) is then that standard linear time series analyses can be applied to these smooth aggregated time series to gage their first order stochastic properties. The main result of this paper is to challenge or qualify this presumption. We show that while it holds in the limit, convergence (as we aggregate) is extremely slow, especially (and paradoxically) when idiosyncratic shocks are large. Moreover, we show that away from this limit, the bias is systematic and it always represents an aggregate time series whose response to aggregate shocks is faster than the true response.

On the constructive side, we show how to use microeconomic data to correct the bias, and demonstrate with a couple of applications the usefulness of this approach. In particular, we show thats the bias can account for the persistence-gap behind Bils and Klenow's (2004) rejection of the Calvo model, and that the difference in the speed with which inflation responds to sectoral and aggregate shocks (Boivin et al 2009; Mackoviak et al 2009) disappears once we correct for the missing persistence bias.

References

[1] Ash, Robert B. and Melvin F. Gardner, *Topics in Stochastic Processes*, New York: Academic Press, 1975.

[2] Bils, Mark, “Do Higher Prices for New Goods Reflect Quality Growth or Inflation,” *The Quarterly Journal of Economics*, **124**(2), May 2009, 637–675.

[3] Bils, Mark and Peter J. Klenow, “Some Evidence on the Importance of Sticky Prices,” *J. of Political Economy*, **112**, 2004, 947–985.

[4] Bils, Mark, Pete Klenow, and Ben Malin, “Reset Price Inflation and the Impact of Monetary Policy Shocks”, *American Economic Review*, **102** (2), October 2012, 2798–2825.

[5] Boivin, Jean, Marc P. Giannoni, and Illian Mihov, “Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data", *American Economic Review*, **102** (2), March 2009, 350–384.

[6] Caballero, Ricardo J., Eduardo M.R.A. Engel, “Price stickiness in *S*s models: New Interpretations of old results”, *Journal of Monetary Economics*, **12**, 2007, 100–121.

[7] Caballero, Ricardo J., Eduardo M.R.A. Engel, and John C. Haltiwanger, “Plant-Level Adjustment and Aggregate Investment Dynamics”, *Brookings Papers on Economic Activity*, 1995 (2), 1–39.

[8] Caballero, Ricardo J., Eduardo M.R.A. Engel, and John C. Haltiwanger, “Aggregate Employment Dynamics: Building from Microeconomic Evidence”, *American Economic Review*, **87** (1), March 1997, 115–137.

[9] Calvo, Guillermo, “Staggered Prices in a Utility-Maximizing Framework,” *Journal of Monetary Economics*, **12**, 1983, 383–398.

[10] Carlsson, Mikael, and Oskar Nordstrom Skans. “Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost." *American Economic Review*, 102(4), June 2012: 1571-95.

[11] Dynan, Karen E., “Habit Formation in Consumer Preferences: Evidence from Panel Data,” *American Economic Review*, **90**(3), June 2000, 391–406.

[12] Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo, “Reference Prices, Costs and Nominal Rigidities", *American Economic Review*, **101** (1), February 2011, 234–262.

[13] Engel, Eduardo M.R.A., “A Unified Approach to the Study of Sums, Products, Time-Aggregation and other Functions of ARMA Processes”, *Journal Time Series Analysis*, **5**, 1984, 159–171.

[14] Gertler, Mark and John Leahy, “A Phillips Curve with an Ss Foundation,” *J. of Political Economy*, **116**, 2008, 533–572.

[15] Goodfriend, Marvin, “Interest-rate Smoothing and Price Level Trend Stationarity,” *Journal of Monetary Economics*, **19**, 1987, 335–348.

[16] Hamilton, James *Time Series Analysis*, Princton University Press, 1994.

[17] Holt, Charles, Franco Modigliani, John Muth and Herbert A. Simon, *Planning Production Inventories, and Work Force*, Prentice-Hall, 1960.

[18] Jorda, Oscar, “Random Time Aggregation in Partial Adjustment Models," *Journal of Business and Economic Statistics*, **7**(3), July 1999, 382–396.

[19] Kehoe, Patrick J., and Vrigiliu Midrigan, “Prices are Sticky After All", *Federal Reserve Bank of Minneapolis Research Department Staff Report*, **413**, June 2012.

[20] Klenow, Pete, and Oleksiy Kryvtsov, “State-Dependent or Time-Dependent Price: Does it Matter for Recent U.S. Inflation", *Quarterly Journal of Economics 123*, 2008, 863–904.

[21] Le Bihan, Herve, and Julien Matheron, “Price Stickiness and sectoral Inflation Persistence: Additional Evidence," *Journal of Money, Credit and Banking*, **44** (7), October 2012, 1427–1422 .

[22] Mackowiak, Bartosz, Emanuel Moench and Mirko Wiederholt, “Sectoral price data and models of price setting,” *J. of Monetary Economics*, **56** October 2009, S78–S99.

[23] Mackowiak, Bartosz and Mirko Wiederholt, “Optimal Sticky Prices under Rational Inattention,” *American Economic Review*, **99** (3), June 2009, 769–803.

[24] Nakamura, Emi, and Jon Steinsson, “Five facts about prices: A reevaluation of menu cost models", *Quarterly Journal of Economics 123*, 2008, 147–180.

[25] Majd, Saman and Robert S. Pindyck, “Time to Build, Option Value, and Investment Decisions,” *Journal of Financial Economics*, **18**, March 1987, 7–27.

[26] Rotemberg, Julio J., “The New Keynesian Microfoundations,” in O. Blanchard and S. Fischer (eds), NBER Macroeconomics Annual, 1987, 69–104.

[27] Sack, Brian, “Uncertainty, Learning, and Gradual Monetary Policy,” Federal Reserve Board Finance and Economics Discussion Series Paper 34, August 1998.

[28] Sargent, Thomas J., “Estimation of Dynamic Labor Demand Schedules under Rational Expectations,” *Journal of Political Economy*, **86**, 1978, 1009–1044.

[29] Sims, Christopher, “Output and Labour Input inManufacturing,” *Brookings Papers on Economic Activity*, 1974, No. 3, 695–735.

[30] Stock, Jim, H., and Mark W. Watson, “Has the Business Cycle Changed and Why?", *NBER Macroeconomics Annual*, 2002, Vol. 17, 159–218.

[31] Tinsley, Peter A., "A Variable Adjustment Model of Labour Demand," *International Economic Review*, 1971, Vol. 12(3), 482–510.

[32] Woodford, Michael, "Optimal Monetary Policy Inertia," NBER WP # 7261, July 1999.

[33] Woodford, Michal, *Interest and prices: Foundations of a theory of monetary policy*, New York: Cambridge University Press, 2005.

APPENDIX

A Additional Bias Correction Methods

In the main text we studied an approach to correct for missing persistence bias using a proxy for y^*, this is the approach we used in Section 4. Here we provide two additional approaches.

A.1 ARMA Correction

The second correction we propose is based on a simple ARMA representation for Δy_t^N.

Proposition 3 (ARMA Representation)

Consider the assumptions and notation of Proposition 1. We then have that Δy_t^N follows the following ARMA(1,1) process:

$$\Delta y_t^N = \rho \Delta y_{t-1}^N + (1-\rho)[\varepsilon_t - \theta \varepsilon_{t-1}], \tag{24}$$

where ε_t is an i.i.d. innovation process and $\theta = (S - \sqrt{S^2 - 4})/2 > 0$ with $S = [2 + (1-\rho^2)(K-1)]/\rho$.[26]

Proof See Appendix C. ∎

Using (24) to write Δy_t^N as an infinite moving average shows that its impulse response to ε-shocks satisfies:

$$I_k = \begin{cases} 1-\rho & \text{if } k = 0 \\ (1-\rho)(\rho-\theta)\rho^{k-1} & \text{if } k \geq 1. \end{cases}$$

Yet this is not the impulse response to the aggregate shock v_t^A, because ε_t in (24) is not v_t^A. As in section 3.3.1, the innovation of the Wold representation is not the innovation of economic interest. The derivation of the impulse response from section 3.3.1 for the case where $N = 1$ carries over to the case with $N > 1$ and the true impulse response is equal to $(1-\rho)\rho^k$, that is, it corresponds to the case where $\theta = 0$ in (24).

This suggests a straightforward approach to estimating the adjustment speed parameter, ρ: Estimate an ARMA(1,1) process (24) and read off the estimate of ρ (and the true impulse response) from the estimated AR-coefficient. That is, first estimate an ARMA model, next drop the MA polynomial and then make inferences about the implied dynamics using only the AR polynomial.

This approach runs into two difficulties when applied in practice. First, for small values of N we have that Δy_t^N is close to an i.i.d. process which means that θ and ρ will be similar. It is well known that estimating an ARMA process with similar roots in the AR and MA polynomials leads to imprecise estimates, resulting in an imprecise estimate for the parameter of interest, ρ.

Second, to apply this approach in a more general setting like the one described by equation (1) in Section 2, the researcher will need to estimate a time-series model with a complex web of AR and MA polynomials and then "drop" the MA polynomial before making inference about the implied dynamics. This strategy is likely to be sensitive to the model specification, for example, the number of lags in the AR-polynomial $b(L)$ in the case of (1).

[26]Scaling the right hand side term by $(1-\rho)$ is inoccuous but useful in what follows.

A.2 Instrumental Variables

Equation (24) in Proposition 1 suggests that lagged values of Δy and Δy^* (or components thereof) may be valid instruments to estimate ρ in a regression of the form

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + e_t.$$

More precisely, if $v_t = \Delta y_t^{*N}$, then Δy_{t-k} and Δy_{t-k}^{*N} will be valid instruments for $k \geq 2$. Yet things are a bit more complicated, since $v_t = \Delta y_t^{*N}$ holds only for $N = \infty$. As shown in the following proposition, the set of valid instruments is larger than suggested above and also includes Δy_{t-1}^{*N}.

Proposition 4 (Instrumental Variables)

With the same notation and assumptions as in Proposition 1, we will have that Δy_{t-k}^N, $k \geq 2$ *and* Δy_{t-j}^{*N}, $j \geq 1$ *are valid instruments when estimating* ρ *from*

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + e_t.$$

By contrast, Δy_{t-1}^N *is not a valid instrument.*

Proof See Appendix C. ∎

B Extensions

B.1 State-dependent Models

The intuition we provided in Section 3 for the missing persistence bias is based on two assumptions: adjustment is lumpy and shocks (the Δy^*) are independent across periods. Thus the correlation between Δy_t and Δy_{t-1} for a unit is zero either because the agent did not adjust in one of the periods or because adjustments at different points in time are independent. This intuition does not depend on whether agents' adjustments are determined by an exogenous process (as in the Calvo model considered so far) or state-dependent (as with Ss-type models). That is, Table 2 in Section 3.3.1 continues to be valid when adjustment policies are state-dependent. because in these models we also have that shocks in non-overlapping time periods are independent when y^* follows a random walk.[27]

Thus the main ingredient for the missing persistence bias is valid both for models with constant and state-dependent adjustment hazards, all that matters is that consecutive adjustments are uncorrelated. Of course, the statistics of interest will be different across both types of models, in particular, the adjustment cost structure is likely to involve more parameters than the sufficient statistics ρ we have worked with so far. Yet the main message remains. For example, when using simulated methods of moments or indirect inference to calibrate or estimate parameters for a DSGE model, using the correct number of agents is important, since otherwise the parameters that are obtained are likely to be biased.

[27] Jorda (1997) provides a general characterization of these models in terms of random point processes (processes with highly localized data distributed randomly in time).

B.2 Relaxing the i.i.d. Assumption

In Section 3 we have assumed that Δy^* is i.i.d. Even though this assumption is a good approximation in many settings (nominal output follows a random walk in Woodford [2003, sect. 3.2], nominal marginal costs follow a random walk in Bils and Klenow [2004]) it is worth exploring what happens when we relax this assumption. When doing so, the cross correlations between contiguous adjustments are no longer zero, but the missing persistence bias typically remains.

We consider first the case where both components of Δy^*, v_t^A and v_{it}^I, follow AR(1) processes with the same first-order autocorrelation ϕ. The case we considered in the main text corresponds to $\phi = 0$. We show in Appendix E that, with a continuum of agents, Δy_t^∞ follows the following stationary ARMA(2,1) process:

$$\Delta y_t^\infty = (\rho+\phi)\Delta y_{t-1}^\infty - \rho\phi\Delta y_{t-2}^\infty + \varepsilon_t - \beta\rho\phi\varepsilon_{t-1},$$

with ε_t proportional to v_t^A and β denoting the agent's discount factor.[28]

Table 7: SLOW CONVERGENCE

Estimated Half-Life and Expected Response Time Δy^* follows an AR(1)

	Effective number of agents (N)						
ϕ	100	400	1,000	4,000	10,000	40,000	True
0	0.252	0.466	0.769	1.724	2.639	3.794	4.596
0.1	0.246	0.440	0.723	1.683	2.659	3.841	4.615
0.2	0.296	0.426	0.686	1.671	2.646	3.852	4.644
0.3	0.379	0.459	0.661	1.615	2.651	3.882	4.690
0.4	0.529	0.564	0.662	1.589	2.697	3.993	4.764
0.5	0.751	0.767	0.801	1.416	2.704	4.064	4.887
0	0.068	0.292	0.684	2.021	3.329	4.988	6.143
0.1	0.069	0.247	0.587	1.932	3.339	5.045	6.160
0.2	0.139	0.246	0.522	1.874	3.290	5.039	6.186
0.3	0.277	0.332	0.509	1.745	3.251	5.050	6.225
0.4	0.514	0.533	0.596	1.661	3.255	5.158	6.288
0.5	0.865	0.870	0.885	1.424	3.183	5.177	6.393

First six rows report the average estimate of the half-life of a shock. The parameter ρ is estimated via maximum likelihood from $(\Delta y_t^N - \phi\Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi\Delta y_{t-2}^N) + e_t - \beta\phi\rho e_{t-1}$ with β and ϕ known. The estimated half-life is obtained by finding k that solves $\sum_{j=0}^k d_k = \frac{1}{2}\sum_{j=0}^\infty d_k$, where $\Delta y_t^N = \sum_{k\geq 0}\psi_k v_{t-k}$ is the (infinite) MA representation of Δy_t^N assumed by the researcher. Estimates based on 100 simulations of length 1,000 each. Rows 7-12 are analogous to rows 1-6 with expected response time instead of estimated half-life. The expected response time is calculated from $(\phi+\rho-2\phi\rho)/(1-\phi-\rho+\rho\phi) - \beta\rho\phi/(1-\beta\rho\phi)$ (see Appendix D). Parameters (monthly pricing data): $\rho = 0.86$, $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$, $\beta = 0.96^{1/12}$.

[28]With the notation of Section 2 we have $b(L) = (1-\phi L)/(1-\beta\rho\phi L)$.

Table 7 shows the measures of speed of convergence considered in Table 1, for the case of prices, once the i.i.d. assumption is relaxed. The first half of the table reports the estimated half-life of a shock, the second half the expected response time. The reported estimates assume that the researcher not only is aware that Δy^* is not i.i.d. but also knows the exact value of the first order autocorrelation, ϕ, as well as β, and estimates ρ via maximum likelihood from

$$(\Delta y_t^N - \phi \Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi \Delta y_{t-2}^N) + e_t - \beta\phi\rho e_{t-1}.$$

The only source of bias is that the researcher ignores the fact that because the actual aggregate considers a finite number of agents, using the linear specification valid for an infinite number of agents will bias the estimated speed of adjustment upwards.[29]

It follows from Table 7 that the bias is generally larger when the Δy^* are correlated than in the i.i.d. case, even though the increase in the bias is small. For example, for $N = 10,000$, the estimated half-life is biased downward by 44.7% when $\phi = 0.5$ as compared with 42.6% when $\phi = 0$. Similarly, the bias for the corresponding expected response times are 45.8 and 50.2%, respectively.

In Section 3 we assumed that y^* is not stationary, we consider next the stationary case. We assume that both the aggregate and idiosyncratic components of y_{it}^* follow stationary AR(1) processes with the same first-order autocorrelation ϕ, in previous sections we assumed $\phi = 1$. The innovations for these processes are the v_t^A and v_{it}^I, respectively. The remaining assumptions remain unchanged.

It follows from Appendix E that, with a continuum of agents, y_t^∞ follows the following stationary AR(2) process:

$$y_t^\infty = (\rho + \phi) y_{t-1}^\infty - \rho\phi y_{t-2}^\infty + \varepsilon_t,$$

with ε_t proportional to v_t^A.

Table 8 revisits Table 1, for annual investment data, this time assuming y^* follows an AR(1) process instead of a random walk. We consider investment, instead of prices as we did in Table 7, because the stationarity assumption for y^* is more reasonable in the case of investment.[30]

Table 8 reports the estimated fraction of adjusting firms, not the estimated half-life or the expected response time. The reason for reporting a persistence measure different from those reported earlier is that when y is stationary the half-life and expected response time for Δy become infinite.[31] Reported estimates assume the researcher knows the value of ϕ in the AR(1) process but believes $N = \infty$, and therefore estimates ρ via OLS from

$$y_t^N - \phi y_{t-1}^N = \rho(y_{t-1}^N - \phi y_{t-2}^N) + e_t. \tag{25}$$

Table 8 shows that the bias is still present when $\phi < 1$ but decreases as ϕ becomes smaller. We show in Appendix F that there is no bias when $\phi = 0$. Because the parameters in Table 8 correspond to annual investment data, the first order autocorrelation parameter ϕ is likely to be around 0.8, suggesting the bias will be large. For example, for $N = 1,000$ (which corresponds roughly to the effective number of firms for the U.S. non-farm business sector) and $\phi = 0.8$, the researcher

[29] Simulations show that the bias disappears if we estimate $(\Delta y_t^N - \phi \Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi \Delta y_{t-2}^N) + e_t - \gamma_1 e_{t-1} - \gamma_2 e_{t-2}$ with no constraints on γ_1 and γ_2. This suggests that the random walk assumption can be relaxed in Proposition 3. We thank Juan Daniel Díaz for this insight.

[30] Nonetheless, results are qualitatively similar if we work with prices.

[31] Also, if we report the half-life and expected response time for y instead of Δy, these persistence measures will be finite but cannot be meaningfully compared with the measures in Table 1 because the latter do not converge to the former when ϕ tends to one.

Table 8: SLOW CONVERGENCE

Estimated Fraction of Adjusters, $1-\rho$, when y^* follows an AR(1)

	Effective number of agents (N)						
ϕ	100	400	1,000	4,000	10,000	40,000	True
0.6	0.493	0.374	0.287	0.198	0.172	0.158	0.150
0.7	0.599	0.448	0.328	0.210	0.177	0.158	0.150
0.8	0.712	0.533	0.385	0.231	0.186	0.161	0.150
0.9	0.843	0.646	0.469	0.269	0.205	0.169	0.150
1.0	0.982	0.856	0.697	0.410	0.279	0.188	0.150

Parameter ρ estimated based on (25), 100 simulations with series of length 1,000. Parameters (annual investment data): $\rho = 0.85$, $\mu_A = 0.12$, $\sigma_A = 0.056$, $\sigma_I = 0.5$, $\beta = 0.96$.

concludes, on average, that 38.5% of firms adjust in any given year, when the true value is 15%.

B.3 Strategic Complementarities

Under the Technical Assumptions from Section 2, agents' decision variables are neither strategic complements nor strategic substitutes. This may not be a reasonable assumption. For example, in the pricing literature many authors have argued that strategic complementarities are a central element to match persistence suggested by VAR evidence.

This motivates considering the case where the y^* are strategic complements. Following Woodford (2003, section 3.2) we assume that log-nominal income follows a random walk with innovations ε_t. Aggregate inflation, π_t, then follows an AR(1) process

$$\pi_t = \phi\pi_{t-1} + (1-\phi)\varepsilon_t$$

with $\phi > \rho$ when prices are strategic complements. In line with the strategic complementarity parameters advocated by Woodford, we assume $\phi = 0.944$. The true half-life of shocks increases from 4.6 to 12.1 months and the expected response time from 6.1 to 16.9 months.

Under these assumptions, $\Delta \log p_t^*$ follows the following ARMA(1,1) process:

$$\Delta \log p_t^* = \phi \Delta \log p_{t-1}^* + c(\varepsilon_t - \rho\varepsilon_{t-1}),$$

with $c = (1-\phi)/(1-\rho)$.[32]

The second and fourth rows in Table 9 present the estimated half-life and expected response time, respectively, in this setting. The first and third rows reproduce the values for the case with no strategic complementarities (Table 1). The bias is larger with strategic complementarities: With 10,000 units, which corresponds to approximately the effective number of prices considered when calculating the CPI, the estimated half-life is one-third of its true value, compared with 60 percent of its true value in the case with no complementarities.

[32] In the notation of Section 2 we have $b(L) = (1-\phi L)/(1-\rho L)$.

Table 9: SLOW CONVERGENCE AND STRATEGIC COMPLEMENTARITIES

Estimated Half-Life with Strategic Complementarities

ρ	ϕ	Effective number of agents (N)						
		100	400	1,000	4,000	10,000	40,000	∞
0.8600	0.8600	0.257	0.464	0.767	1.744	**2.699**	3.886	4.595
0.8600	0.9442	0.268	0.484	0.826	2.170	**4.016**	7.638	12.067
0.8600	0.8600	0.072	0.290	0.681	2.049	**3.415**	5.121	6.142
0.8600	0.9442	0.081	0.314	0.761	2.657	**5.308**	10.527	16.914

First two rows show the estimated half-life. The half-life is calculated from $-\log 2/\log\hat{\rho}_\infty$ with $\hat{\rho}_\infty \equiv \text{plim}_{T\to\infty}\hat{\rho}$ when $\rho = \phi$ and $\hat{\rho}$ estimated from (7) with 100 simulations of length 1000 when $\phi > \rho$. Rows 3-4 show results when the expected response time (ERT) is the measure of persistence. For an AR(1), ERT is defined as $\hat{\rho}_\infty/(1-\hat{\rho}_\infty)$. Parameters: $\rho = 0.86$, $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$. Numbers in boldface correspond to the effective number of units for U.S. CPI.

B.4 Adding smooth adjustment

Suppose now that in addition to the infrequent adjustment pattern described above, once adjustment takes place, it is only gradual. Such behavior is observed, for example, when there is a time-to-build feature in investment (e.g., Majd and Pindyck (1987)) or when policy is designed to exhibit inertia (e.g., Goodfriend (1987), Sack (1998), or Woodford (1999)). Our main result here is that the econometrician estimating a linear ARMA process —a Calvo model with additional serial correlation— will only be able to extract the gradual adjustment component but not the source of sluggishness from the infrequent adjustment component. That is, again, the estimated speed of adjustment will be too fast, for exactly the same reason as in the simpler model.

Let us modify our basic model so that equation (2) now applies for a new variable $\tilde{y}_t$ in place of y_t, with $\Delta\tilde{y}_t$ representing the *desired* adjustment of the variable that concerns us, Δy_t. This adjustment takes place only gradually, for example, because of a time-to-build component. We capture this pattern with the process:

$$\Delta y_t = \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + (1 - \sum_{k=1}^{K} \phi_k)\Delta\tilde{y}_t. \tag{26}$$

Now there are two sources of sluggishness in the transmission of shocks, Δy_t^*, to the observed variable, Δy_t. First, the agent only acts intermittently, accumulating shocks in periods with no adjustment. Second, when the agent adjusts, it does so only gradually.

By analogy with the simpler model, suppose the econometrician approximates the lumpy component of the more general model by:

$$\Delta\tilde{y}_t = \rho\Delta\tilde{y}_{t-1} + v_t. \tag{27}$$

Replacing (27) into (26), yields the following linear equation in terms of the observable, Δy_t:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t, \tag{28}$$

with

$$\begin{aligned} a_1 &= \phi_1 + \rho, \\ a_k &= \phi_k - \rho\phi_{k-1}, \qquad k = 2, ..., K, \\ a_{K+1} &= -\rho\phi_K, \end{aligned} \tag{29}$$

and $\varepsilon_t \equiv (1-\rho)(1-\sum_{k=1}^K \phi_k)\Delta y_t^*$.

By analogy to the simpler model, we now show that the econometrician will miss the source of persistence stemming from ρ.

Proposition 5 (Omitted Source of Sluggishness)

Let all the assumptions in Proposition 1 hold, with $\tilde{y}$ in the role of y. Also assume that (26) applies, with all roots of the polynomial $1-\sum_{k=1}^K \phi_k z^k$ outside the unit circle. Let $\hat{a}_k$, $k = 1, ..., K+1$ denote the OLS estimates of equation (28).

Then:

$$\begin{aligned} \text{plim}_{T\to\infty} \hat{a}_k &= \phi_k, \qquad k = 1, ..., K, \\ \text{plim}_{T\to\infty} \hat{a}_{K+1} &= 0. \end{aligned} \tag{30}$$

Proof See Appendix C. ∎

Comparing (29) and (30) we see that the proposition simply reflects the fact that the (implicit) estimate of ρ is zero.

C Proof of Propositions

Proof of Proposition 1

In this appendix we prove Proposition 1. The proof uses an auxiliary variable, x_{it}, equal to how much unit i adjusts in period t if it adjusts that period (that is, the value of Δy_{it} conditional on adjustment). Because of the Technical Assumptions, x_{it} equals the unit's accumulated shocks since it last adjusted. The following dynamic dynamic definition of x_{it} is what we use in the proof:

$$x_{i,t+1} = (1-\xi_{it})x_{it} + \Delta y^*_{i,t+1}, \tag{31}$$

$$\Delta y_{it} = \xi_{it} x_{it}. \tag{32}$$

In what follows, subindices i and j denote *different* units.

We first derive the following unconditional expectations:

$$\begin{aligned}
\mathrm{E}x_{it} &= \frac{\mu_A}{1-\rho}, &(33)\\
\mathrm{E}[\Delta y_{it}] &= \mu_A, &(34)\\
\mathrm{E}[\Delta y_t^N] &= \mu_A, &(35)\\
\mathrm{E}[x_{it}x_{jt}] &= \frac{1}{1-\rho^2}\left[\sigma_A^2+\frac{1+\rho}{1-\rho}\mu_A^2\right], &(36)\\
\mathrm{E}[x_{it}^2] &= \frac{1}{1-\rho}\left[\sigma_A^2+\sigma_I^2+\frac{1+\rho}{1-\rho}\mu_A^2\right]. &(37)
\end{aligned}$$

From (31) and the Technical Assumption in the main text we have:

$$\mathrm{E}x_{i,t+1} = \rho\mathrm{E}x_{it} + \mu_A.$$

The above expression leads to (33) once we note that the stationarity of x_{it} implies $\mathrm{E}x_{i,t+1} = \mathrm{E}x_{it}$.

Equation (34) follows from (33) and Technical Assumption 3. Equation (35) follows directly from (34).

To derive (36), we note that, from (31)

$$\begin{aligned}
\mathrm{E}[x_{i,t+1}x_{j,t+1}] &= \mathrm{E}[\{(1-\xi_{it})x_{it}+\Delta y^*_{i,t+1}\}\{(1-\xi_{jt})x_{jt}+\Delta y^*_{j,t+1}\}]\\
&= \mathrm{E}[(1-\xi_{it})x_{it}(1-\xi_{jt})x_{jt}] + \mathrm{E}[\Delta y^*_{i,t+1}(1-\xi_{jt})x_{jt}]\\
&\qquad + \mathrm{E}[(1-\xi_{it})x_{it}\Delta y^*_{j,t+1}] + \mathrm{E}[\Delta y^*_{i,t+1}\Delta y^*_{j,t+1}]\\
&= \rho^2\mathrm{E}[x_{it}x_{jt}] + 2\frac{\rho}{1-\rho}\mu_A^2 + (\mu_A^2+\sigma_A^2),
\end{aligned}$$

where we used the Technical Assumptions, (33) and $i \neq j$. Noting that $x_{it}x_{jt}$ is stationary and therefore $\mathrm{E}[x_{it}x_{jt}] = \mathrm{E}[x_{i,t-1}x_{j,t-1}]$, the above expression leads to (36).

Finally, to prove (37), we note that, from (31) we have

$$\begin{aligned}
\mathrm{E}[x_{i,t+1}^2] &= \mathrm{E}[(1-\xi_{it})x_{it}^2] + 2\mathrm{E}(1-\xi_{it})x_{it}\Delta y^*_{i,t+1}] + \mathrm{E}[(\Delta y^*_{i,t+1})^2]\\
&= \rho\mathrm{E}[x_{it}^2] + 2\frac{\rho}{1-\rho}\mu_A^2 + (\sigma_A^2+\sigma_I^2+\mu_A^2),
\end{aligned}$$

where we used that $(1-\xi_{it})^2 = 1-\xi_{it}$, (33) and the Technical Assumptions. Stationarity of x_{it} (and therefore x_{it}^2) and some simple algebra complete the proof.

Next we use the five unconditional expectations derived above to obtain the four expressions in the second row of Table 3. The expression for the OLS estimate $\hat{\rho}$ in (8) then follows from tedious but otherwise straightforward algebra.

We have:

$$\begin{aligned}
&\mathrm{Cov}(\Delta y_{i,t+1},\Delta y_{it}) = \mathrm{E}[\Delta y_{i,t+1}\Delta y_{it}] - \mu_A^2 = \mathrm{E}[\xi_{i,t+1}x_{i,t+1}\xi_{it}x_{it}] - \mu_A^2 = (1-\rho)\mathrm{E}[x_{i,t+1}\xi_{it}x_{it}] - \mu_A^2\\
&\quad = (1-\rho)\mathrm{E}[\{(1-\xi_{it})x_{it}+\Delta y^*_{i,t+1}\}\xi_{it}x_{it}] - \mu_A^2 = (1-\rho)\mathrm{E}[\{(1-\xi_{it})\xi_{it}x_{it}^2] + (1-\rho)\mathrm{E}[\Delta y^*_{i,t+1}\xi_{it}x_{it}] - \mu_A^2\\
&\quad = (1-\rho)\times 0 + (1-\rho)\mu_A^2 - \mu_A^2 = -\rho\mu_A^2,
\end{aligned}$$

where in the crucial step we used that $(1-\xi_{it})\xi_{it}$ always equals zero.

We also have the cross-covariance terms ($i \neq j$):

$$\begin{aligned}
\mathrm{Cov}(\Delta y_{i,t+1}, \Delta y_{jt}) &= \mathrm{E}[\xi_{i,t+1}x_{i,t+1}\xi_{jt}x_{jt}] - \mu_A^2 = (1-\rho)\mathrm{E}[x_{i,t+1}\xi_{jt}x_{jt}] - \mu_A^2 \\
&= (1-\rho)\mathrm{E}[\{(1-\xi_{it})x_{it} + \Delta y^*_{i,t+1}\}\xi_{jt}x_{jt}] - \mu_A^2 = \rho(1-\rho)^2\mathrm{E}[x_{it}x_{jt}] + (1-\rho)\mu_A^2 - \mu_A^2 = \frac{1-\rho}{1+\rho}\rho\sigma_A^2. \\
\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt}) &= \mathrm{E}[\xi_{it}x_{it}\xi_{jt}x_{jt}] - \mu_A^2 = (1-\rho)^2\mathrm{E}[x_{it}x_{jt}] - \mu_A^2 = \frac{1-\rho}{1+\rho}\sigma_A^2.
\end{aligned}$$

Finally, the variance term is obtained as follows:

$$\mathrm{Var}(\Delta y_{it}) = \mathrm{E}[\xi_{it}^2 x_{it}^2] - \mu_A^2 = \mathrm{E}[\xi_{it}x_{it}^2] - \mu_A^2 = (1-\rho)\mathrm{E}[x_{it}^2] - \mu_A^2 = \sigma_A^2 + \sigma_I^2 + \frac{2\rho}{1-\rho}\mu_A^2. \quad \blacksquare$$

Proof of Proposition 2

Part (i) follows trivially from Proposition 1 and the fact that both regressors are uncorrelated. To prove (ii) we first note that:

$$\mathrm{plim}_{T\to\infty}\hat{b}_1 = \frac{\mathrm{Cov}(\Delta y_t - \Delta y_{t-1}\,,\, \Delta y_t^* - \Delta y_{t-1})}{\mathrm{Var}(\Delta y_t^* - \Delta y_{t-1})}.$$

We therefore need expressions for $\mathrm{Cov}(\Delta y_t^N, \Delta y_t^{N*})$, $\mathrm{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)$ and $\mathrm{Var}(\Delta y_t^N)$.
We have

$$\mathrm{Cov}(\Delta y_t^N, \Delta y_t^{N*}) = \frac{1}{N}\mathrm{Cov}(\Delta y_{it}, \Delta y_{it}^*) + \left(1 - \frac{1}{N}\right)\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt}).$$

Both covariances on the r.h.s. are calculated using (31), yielding $\sigma_A^2 + \sigma_I^2$ and σ_A^2, respectively. Expressions for $\mathrm{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)$ and $\mathrm{Var}(\Delta y_t^N)$ are obtained using an analogous decomposition and the covariances and variances from Table 3. We have all the terms for the expression above for $\hat{b}_1$, the remainder of the proof is some tedious but otherwise straightforward algebra. $\blacksquare$

Proof of Proposition 3

To prove that Δy_t^N follows an ARMA(1,1) process with autoregressive coefficient ρ, it suffices to show that the process's autocorrelation function, γ_k, satisfies:[33]

$$\gamma_k = \rho\gamma_{k-1}, \qquad k \geq 2. \tag{38}$$

We prove this next and derive the moving average parameter θ by finding the unique θ within the unit circle that equates the first-order autocorrelation of this process, which by Proposition 1 is given by (8), with the following well known expression for the first order autocorrelation of an ARMA(1,1) process:

$$\gamma_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}.$$

Proving that θ tends to zero as N tends to infinity is straightforward.

[33]Here we are using Theorem 1 in Engel (1984) characterizing ARMA processes in terms of difference equations satisfied by their autocorrelation function.

We have:

$$\begin{aligned}
\mathrm{E}[\Delta y^N_{t+k}\Delta y^N_t] &= \sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[\xi_{i,t+k}x_{i,t+k}\xi_{jt}x_{jt}] \\
&= (1-\rho)\sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[x_{i,t+k}\xi_{jt}x_{jt}] \\
&= (1-\rho)\sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[\{(1-\xi_{i,t+k-1})x_{i,t+k-1}+\Delta y^*_{i,t+k}\}\xi_{jt}x_{jt}] \\
&= (1-\rho)\rho\sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[x_{i,t+k-1}\xi_{jt}x_{jt}] \;+\; (1-\rho)\mu_A\sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[\xi_{jt}x_{jt}] \\
&= \rho\sum_{i=1}^n\sum_{j=1}^n w_i w_j \mathrm{E}[\xi_{i,t+k-1}x_{i,t+k-1}\xi_{jt}x_{jt}] \;+\; (1-\rho)\mu_A^2 \\
&= \rho\mathrm{E}[\Delta y^N_{t+k-1}\Delta y^N_t] + (1-\rho)\mu_A^2,
\end{aligned}$$

where in the fourth step we assumed $k \geq 2$, since we used that $\xi_{i,t+k-1}$ and ξ_{jt} are independent even when $i = j$. Noting that $\gamma_k = (\mathrm{E}[\Delta y^N_{t+k}\Delta y^N_t] - \mu_A^2)/\mathrm{Var}(\Delta y_t)$ and using the above identity yields (38) and concludes the proof. ∎

Proof of Proposition 4

We have:

$$\Delta y^N_t = \sum_i w_i\xi_{it}x_{it} = \sum_i w_i\xi_{it}(y^*_{it}-y_{i,t-1}) = \sum_i w_i(1-\rho)(y^*_{it}-y_{i,t-1})+\sum_i w_i(\xi_{it}-1+\rho)(y^*_{it}-y_{i,t-1}).$$

Similarly

$$\Delta y^N_{t-1} = \sum_i w_i(1-\rho)(y^*_{i,t-1}-y_{i,t-2})+\sum_i w_i(\xi_{i,t-1}-1+\rho)(y^*_{i,t-1}-y_{i,t-2}).$$

Subtracting the latter from the former and rearranging terms yields

$$\Delta y^N_t = \rho\Delta y^N_{t-1} + (1-\rho)\Delta y^{*N}_t + \epsilon^N_t \tag{39}$$

with

$$\epsilon^N_t = \sum_i w_i\left[(\xi_{it}-1+\rho)(y^*_{it}-y_{i,t-1}) - (\xi_{i,t-1}-1+\rho)(y^*_{i,t-1}-y_{i,t-2})\right]. \tag{40}$$

The extra term ϵ^N_t on the r.h.s. of (40) explains why Δy^N_{t-1} is not a valid instrument: Δy^N_{t-1} is correlated with ϵ^N_t because both include $\xi_{i,t-1}$ terms. Of course, ϵ^N_t tends to zero as N tends to infinity: its mean is zero and a calculation using many of the expressions derived in the proof of Proposition 1 shows that

$$\mathrm{Var}(\epsilon_t) = \frac{2\rho}{N}\left[\sigma_A^2+\sigma_I^2+\frac{1+\rho}{1-\rho}\mu_A^2\right].$$

It follows from (39), (40) and Technical Assumption 3 that ϵ_t is uncorrelated with Δy^*_s, for all s, which implies that Δy^*_{t-s} is a valid instrument for $s \geq 1$. And since $\Delta y_{i,t-k}$ are uncorrelated with ξ_{it} and $\xi_{i,t-1}$ for $k \geq 2$, we have that lagged values of Δy, with at least two lags, are valid instruments as well. ∎

Proof of Proposition 5

The equation we estimate is:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t, \tag{41}$$

while the true relation is that described by (26) and (27).

It is easy to see that the second term on the right hand side of (26) denoted by w_t in what follows, is uncorrelated with Δy_{t-k}, $k \geq 1$. It follows that estimating (41) is equivalent to estimating (26) with error term

$$w_t == (1 - \sum_{k=1}^{K} \phi_k)\xi_t \sum_{k=0}^{l_t - 1} \Delta y^*_{t-k},$$

and therefore:

$$\text{plim}_{T\to\infty} \hat{a}_k = \begin{cases} \phi_k & \text{if } k = 1, 2, ..., K, \\ 0 & \text{if } k = K+1. \end{cases}$$

This concludes the proof. ∎

D The Expected Response Time Index: τ

We define the expected response time of Δy to Δy^* as:

$$\tau \equiv \frac{\sum_{k\geq 0} k I_k}{\sum_{k\geq 0} I_k}, \tag{42}$$

with

$$I_k \equiv \mathrm{E}_t \left[\frac{\partial \Delta y_{t+k}}{\partial \epsilon_t} \right].$$

Where $\mathrm{E}_t[\cdot]$ denotes expectations conditional on information (that is, values of Δy and Δy^*) known at time t. This index is a weighted sum of the components of the impulse response function, with weights proportional to the number of periods that elapse until the corresponding response is observed. For example, an impulse response with the bulk of its mass at low lags has a small value of τ, since Δy responds relatively fast to shocks.

Lemma A1 (τ for an Infinite MA) *Consider a second order stationary stochastic process*

$$\Delta y_t = \sum_{k\geq 0} \psi_k \epsilon_{t-k},$$

with $\psi_0 = 1$, $\sum_{k\geq 0} \psi_k^2 < \infty$, the ϵ_t's uncorrelated, and ϵ_t uncorrelated with $\Delta y_{t-1}, \Delta y_{t-2}, ...$ Assume that $\Psi(z) \equiv \sum_{k\geq 0} \psi_k z^k$ has all its roots outside the unit disk.

Then:

$$I_k = \psi_k \quad \textit{and} \quad \tau = \frac{\Psi'(1)}{\Psi(1)} = \frac{\sum_{k\geq 1} k\psi_k}{\sum_{k\geq 0} \psi_k}.$$

Proof That $I_k = \psi_k$ is trivial. The expressions for τ then follow from differentiating $\Psi(z)$ and evaluating at $z = 1$. ∎

Proposition A1 (**τ for an ARMA Process**) *Assume Δy_t follows an ARMA(p,q):*

$$\Delta y_t - \sum_{k=1}^{p} \phi_k \Delta y_{t-k} = \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k},$$

where $\Phi(z) \equiv 1 - \sum_{k=1}^{p} \phi_k z^k$ and $\Theta(z) \equiv 1 - \sum_{k=1}^{q} \theta_k z^k$ have all their roots outside the unit disk. The assumptions regarding the ϵ_t's are the same as in Lemma A1.

Define τ as in (42)*. Then:*

$$\tau = \frac{\sum_{k=1}^{p} k\phi_k}{1-\sum_{k=1}^{p} \phi_k} - \frac{\sum_{k=1}^{q} k\theta_k}{1-\sum_{k=1}^{q} \theta_k}.$$

Proof Given the assumptions we have made about the roots of $\Phi(z)$ and $\Theta(z)$, we may write:

$$\Delta y_t = \frac{\Theta(L)}{\Phi(L)} \epsilon_t,$$

where L denotes the lag operator. Applying Lemma A1 with $\Theta(z)/\Phi(z)$ in the role of $\Psi(z)$ we then have:

$$\tau = \frac{\Theta'(1)}{\Theta(1)} - \frac{\Phi'(1)}{\Phi(1)} = \frac{\sum_{k=1}^{p} k\phi_k}{1-\sum_{k=1}^{p} \phi_k} - \frac{\sum_{k=1}^{q} k\theta_q}{1-\sum_{k=1}^{q} \theta_k}. \quad \blacksquare$$

Proposition A2 (**τ for a Lumpy Adjustment Process**) *Consider Δy_t in the simple lumpy adjustment model (12) and τ defined in* (42)*. Then $\tau = \rho/(1-\rho)$.*

Proof $\partial \Delta y_{t+k} / \partial \Delta y_t^*$ is equal to one when the unit adjusts at time $t+k$, not having adjusted between times t and $t+k-1$, and is equal to zero otherwise. Thus:

$$I_k \equiv \mathrm{E}_t \left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*} \right] = \Pr\{\xi_{t+k} = 1 \,,\, \xi_{t+k-1} = \xi_{t+k-2} = \ldots = \xi_t = 0\} = (1-\rho)\rho^k. \tag{43}$$

The expression for τ now follows easily. $\blacksquare$

E Rotemberg's Equivalence Result

Proposition 6 (Rotemberg's Equivalence Result)

Agent i controls y_{it}, $i = 1, \ldots, N$. The aggregate value of y is defined as $y_t^N \equiv \frac{1}{N} \sum_{i=1}^{N} y_{it}$. In every period, the cost of changing y is either infinite (with probability ρ) or zero (with probability $1-\rho$) (Calvo Model). When the agent adjusts, it chooses y_{it} equal to $\tilde{y}_t$ that solves

$$\min_{\tilde{y}_t} \mathrm{E}_t \sum_{k \geq 0} (\beta\rho)^k (y_{t+k}^* - \tilde{y}_t)^2,$$

where β denotes the agent's discount factor and y_t^* denotes an exogenous process.[34] *We then have*

$$\tilde{y}_t = (1-\beta\rho) \sum_{k \geq 0} (\beta\rho)^k \mathrm{E}_t y_{t+k}^*. \tag{44}$$

[34] This formulation can be extended to incorporate idiosyncratic shocks.

It follows that, as N tends to infinity, y_t^∞ satisfies:

$$y_t^\infty = \rho y_{t-1}^\infty + (1-\rho)\tilde{y}_t. \tag{45}$$

Consider next an alternative adjustment technology (Quadratic Adjustment Costs) where in every period agent i choose y_{it} that solves:

$$\min_{y_{it}} \mathrm{E}_t \sum_{k\geq 0} \beta^k [(y_{t+k}^* - y_{it})^2 + c(y_{it} - y_{i,t-1})^2],$$

where $c > 0$ captures the relative importance of quadratic adjustment costs. We then have that there exists $\rho' \in (0,1)$ and $\delta \in (0,1)$ s.t.[35]

$$y_t^\infty = \rho' y_{t-1}^\infty + (1-\rho')\hat{y}_t, \tag{46}$$

with

$$\hat{y}_t = (1-\delta)\sum_{k\geq 0} \delta^k \mathrm{E}_t y_{t+k}^*. \tag{47}$$

Finally, and this is Rotemberg's contribution, a comparison of (44)-(45) *and* (46)-(47) *shows that an econometrician working with aggregate data cannot distinguish between the Calvo model and the Quadratic Adjustment Costs model described above: ρ' plays the role of ρ and δ the role of $\beta\rho$.*

Proof See Rotemberg (1987). ∎

Corollary 1 *Under the assumptions of the Calvo Model in Proposition 6.*

a) *Consider the case where y_t^* follows an AR(1):*

$$y_t^* = \psi y_{t-1}^* + e_t,$$

with $|\psi| < 1$. We then have that $\mathrm{E}_t y_{t+k}^* = \psi^k y_t^*$ and y_t^∞ follows the following AR(2) process:

$$y_t^\infty = (\rho+\psi)y_{t-1}^\infty - \rho\psi y_{t-2}^\infty + \frac{(1-\rho)(1-\beta\rho)}{1-\beta\rho\psi} e_t. \tag{48}$$

b) *Consider the case where Δy_t^* follows an AR(1):*

$$\Delta y_t^* = \phi \Delta y_{t-1}^* + e_t,$$

with $|\phi| < 1$. We then have that

$$\mathrm{E}_t y_{t+k}^* = \frac{\phi(1-\phi^k)}{1-\phi}\Delta y_t^* + y_t^*$$

and Δy_t^∞ follows the following ARMA(2,1) process:

$$\Delta y_t^\infty = (\rho+\phi)\Delta y_{t-1}^\infty - \rho\phi\Delta y_{t-2}^\infty + \frac{1-\rho}{1-\beta\rho\phi}[e_t - \beta\rho\phi e_{t-1}].$$

[35] The expression that follows is equivalent to the partial adjustment formulation:

$$\Delta y_t^\infty = (1-\rho')(\hat{y}_t - y_{t-1}^\infty),$$

Proof Straightforward. ∎

F The case where y^* is i.i.d.

Assume that

$$y^*_{it} = y^{*A}_t + y^{*I}_{it}$$

with y^{*A}_t i.i.d. with mean μ_A and variance σ^2_A and y^{*I}_{it} i.i.d. with zero mean and variance σ^2_I. The y^{*I}_{it} processes are independent across agents and independent from the aggregate shock process y^{*A}_t. The remaining assumptions are the same as in the Technical Assumptions we made in Section 2.

For simplicity we assume $\mu_A = 0$, the case where $\mu_A \neq 0$ just adds a constant to the expressions that follow. Equation (48) then implies that:

$$y^\infty_t = \rho y^\infty_{t-1} + (1-\rho)(1-\beta\rho) y^{*A}_t. \tag{49}$$

We show next that the OLS estimator of ρ in the regression

$$y^\infty_t = \rho y^\infty_{t-1} + e_t \tag{50}$$

provides a consistent estimator of ρ even when N is finite. That is, when the driving processes y^* are i.i.d., there is no missing persistence bias.

Extending the analysis (and notation) from Appendix E to incorporate idiosyncratic shocks, we obtain

$$\tilde{y}_{it} = (1-\beta\rho) y^*_{it}.$$

Using the notation we introduced in Appendix C this implies that

$$y^N_t = \frac{1}{N}\sum_{i=1}^N (1-\xi_{it}) y_{i,t-1} + (1-\beta\rho)\frac{1}{N}\sum_{i=1}^N \xi_{it} y^*_{it}.$$

Following a similar logic to the one we used in the proof of Proposition 4, we can rewrite the above expression as

$$y^N_t = \rho y^N_{t-1} + \varepsilon_t \tag{51}$$

with

$$\varepsilon_t = \frac{1}{N}\sum_{i=1}^N (1-\xi_{it}-\rho) y_{i,t-1} + (1-\beta\rho)\frac{1}{N}\sum_{i=1}^N \xi_{it} y^*_{it}.$$

Even though ε_t differs from the error term in (49), it also is uncorrelated with the regressor y^N_{t-1} which is all we need for $\hat{\rho}$ estimated via OLS from (51) to be a consistent estimator for ρ.

G Reset price inflation and estimation of sectoral shocks

As mentioned in the main text, we use the reset price inflation introduced by Bils, Klenow and Mailn (2012) to estimate sectoral shocks. The discussion of their methodology in this section closely mirrors their own. The only difference is that we estimate reset price inflation at the sectoral level, whereas Bils, Klenow and Malin (2012) focus on the aggregate properties of reset price inflation. The basic idea behind reset price inflation is to make inferences about the underlying shocks using information contained only from observed price changes where the implicit assumption is that when a firm adjusts it is adjusting to its optimal reset price.

Specifically, define $p_{i,t}$ as the log price of item i and time t and define a price change indicator as:

$$I_{i,t} = \begin{cases} 1 & \text{if} \quad p_{i,t} \neq p_{i,t-1} \\ 0 & \text{if} \quad p_{i,t} = p_{i,t-1} \end{cases}$$

For prices that change, the reset price, $p_{i,t}^{\text{reset}}$, is simply the current price. For prices that do not change, we index our estimate of the reset price to the rate of reset price inflation among price changers in the current period.

$$p_{i,t}^{\text{reset}} = \begin{cases} p_{i,t} & I_{i,t} = 1 \\ p_{i,t-1} + \pi_t^{\text{reset}} & I_{i,t} = 0 \end{cases}.$$

Given $p_{i,t-1}^{\text{reset}}$, define reset price inflation, π_t^{reset}, as:

$$\pi_t^{\text{reset}} = \frac{\sum_i \omega_{i,t} \left(p_{i,t} - p_{i,t-1}^{\text{reset}} \right) I_{i,t}}{\sum_i \omega_{i,t} I_{i,t}}.$$

where $\omega_{i,t}$ denote i's relative expenditure weight at time t. Thus reset price inflation is the "inflation rate" conditional on the price changers. With Calvo price setting and assuming that the technical assumptions in Section 3 hold, it is easy to show that reset price inflation reduces to the following formula:[36]

$$\pi_t^{\text{reset}} = \frac{\pi_t - \rho \pi_{t-1}}{(1-\rho)} = \nu_t^A$$

This justifies using the reset price inflation methodology to estimate aggregate shocks. In Appendix G we present simulation results showing that reset price inflation is also a good method to recover the true shock innovations in both more realistic Calvo environments with large idiosyncratic shocks and Ss-type settings.

H Simulation details

H.1 Calibration details

The details of the multi-sector Calvo model calibration are as follows. We calibrate a 66 sector version of the Calvo pricing model. For each sector, we set the average sectoral inflation rate to what is observed in the CPI micro data. We choose the standard deviation of the sectoral inflation rate series, the persistence and standard deviation of the sectoral idiosyncratic shock series (assumed to be an AR(1) in logs) to match the following four moments: the average size of price increases and decreases, the fraction of price changes that are price increases and the standard deviation of the sectoral inflation rate. In the model, the number of firms in each sector is given by the median (across time) number of firms for that sector in the micro BLS data and each firm was simulated for 238 periods, which is the number of periods in the underlying data.

[36]This holds in the limit as the number of price setters becomes large so that the frequencies are exact and the idiosyncratic shocks average out.

Table 10 shows basic descriptive statistics for the simulated model, reported statistics are medians across the 66 sectors, suggesting that the multi-sector Calvo model does a good job matching moments across sectors.

Table 10: DETAILS OF MULTI-SECTOR CALVO CALIBRATION

Calibration results: Basic Statistics

	CPI	Model
Frequency of monthly adjustment:	0.068	0.068
Fraction price changes > 0:	0.669	0.567
Average size of increases (%):	7.997	8.305
Average size of decreases (%):	9.073	8.180
Std of sectoral inflation:	0.004	0.005

H.2 Monte-Carlo Evidence: Do We Recover the True Shock In Practice?

To see if our shock measure was recovering the true aggregate shock, we simulated both a Calvo and an Ss model with the following standard parameter values: the frequency of adjustment = 0.2, $\mu_{\text{agg}} = 0.002$, $\sigma_{\text{agg}} = 0.003$, $\rho_I = 0.97; \sigma_I = 0.04$ (also tried something farther from a random walk: ρ_I =0.7) These economies were simulated for T=300 periods with a burn in of 100 periods. Notice that there are two types of shocks: aggregate shocks that affect everyone and idiosyncratic shocks that are firm specific. In each simulation we ran the following regression:

$$v_t = \alpha + \beta z_t + e_t$$

where v_t is our shock measure (reset price inflation) and z_t is the true shock innovation from each simulation. The level and fit of this regression is informative of how well our shock measure proxies for the true shock. It is an important robustness check because we want to make sure that we can recover an unbiased estimate of the true aggregate shock in a situation where idiosyncratic shocks are realistically large relative to aggregate shocks. The results (averaged across 100 simulations) are comforting and shown below:

Unsurprisingly, the overall fit improves in terms of R^2 as the sample sizes increase. Most importantly, we recover the true innovations in the Calvo case and an affine transformation of the innovations in the Ss case for all sample sizes.

Table 11: DOES RESET PRICE INFLATION RECOVER THE TRUE SHOCKS?

REGRESSION OF ESTIMATED SHOCK ON TRUE SHOCK: RESET PRICE INFLATION

			CALVO			SS	
	NFIRMS	INTERCEPT	SLOPE	R^2	INTERCEPT	SLOPE	R^2
$\rho = .7$	500	-0.00	1.02	0.34	-0.00	3.07	0.41
		(0.00)	(0.08)	(0.04)	(0.00)	(0.19)	(0.04)
	5000	-0.00	1.04	0.76	-0.00	3.05	0.67
		(0.00)	(0.03)	(0.02)	(0.00)	(0.18)	(0.04)
	25000	-0.00	1.04	0.85	-0.00	3.07	0.72
		(0.00)	(0.02)	(0.02)	(0.00)	(0.10)	(0.03)
$\rho = .97$	500	-0.00	0.99	0.07	-0.00	2.97	0.28
		(0.00)	(0.21)	(0.03)	(0.00)	(0.26)	(0.04)
	5000	-0.00	1.02	0.35	-0.00	3.00	0.45
		(0.00)	(0.07)	(0.05)	(0.00)	(0.20)	(0.04)
	25000	-0.00	1.01	0.51	-0.00	3.00	0.48
		(0.00)	(0.06)	(0.04)	(0.00)	(0.22)	(0.03)