MISSING AGGREGATE DYNAMICS:
ON THE SLOW CONVERGENCE OF LUMPY ADJUSTMENT MODELS

David Berger  
Northwestern

Ricardo Caballero  
MIT and NBER

Eduardo Engel  
U. of Chile and NBER

October 13, 2014

Abstract

When applying conventional VAR procedures to aggregate variables with lumpy micro adjustment, estimated persistence of shocks is biased downwards. This is relevant for non-, semi- and structural models in macroeconomics. The extent to which persistence is underestimated 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 aggregate 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 find that the difference in the speed with which inflation responds to sectoral and aggregate shocks 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, Calvo model, Ss model, inflation, sectoral shocks, aggregate shocks, investment, labor demand, sticky prices, biased impulse response functions, vector auto regressions.

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, Eric Swanson, Harald Uhlig, Joe Vavra and seminar participants at Chicago Fed, University of Chicago, FGV (Rio and Sao Paulo), Humboldt Universität, MIT, NBER SI (EFCE) PUC (Chile), Universidad de Chile (CEA and FEN), University of Maryland, University of Paris, University of Pennsylvania, Yale University, NBER EFG Meeting, the 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 70,000 prices, the bias turns out to be large, with the estimated half-life of shocks biased downward by approximately 40%.

We propose three procedures for correcting the bias we highlight in this paper. One estimates an ARMA specification for the aggregate of interest that captures the true underlying dynamics in the AR component. Another uses instrumental variables while the third approach estimates the underlying shocks. We also 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, 2008). 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 (and are both slower than those estimated ignoring the bias).

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 even if 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 speed of adjustment. That is, even though the nonlinear IRF recovers the Rotemberg (1987) result, according to which the aggregate of interest follows an AR(1) 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.

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 its true value 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. Since in practice idiosyncratic uncertainty is many times larger than aggregate uncertainty, we conclude that the problem of missing aggregate dynamics is 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 well the actual persistence of shocks. By contrast, a central theme underlying the results in this paper is that when the variable of interest aggregate over units with lumpy adjustment, using a VAR will underestimate the true persistence of shocks. The shocks inferred

---

\(^2\)The theoretical underpinning for this statement is Wold's representation result, see Ash and Gardner, 1975, for an insightful discussion.
from the VAR estimation differ systematically from the true underlying shocks, 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 aggregates with lumpy microeconomic adjustment, as long as the number of units in the aggregate is infinite. Section 3 presents the missing persistence bias that arises when the number of units considered is finite. This section also provides results establishing the slow convergence to the Rotemberg limit. Section 4 describes three approaches to correct for the bias while Section 5 considers various extensions of the baseline model and shows that the bias is present, and continues being significant. Section 6 studies two detailed applications and Section 7 concludes. Various appendices follow.

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, at some key stage a researcher estimates an equation of the form:

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

where $\Delta y$ 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; $\varepsilon$ 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 $a_k$s 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 'rest value' of $y$). We will have that:

$$\Delta y_{it} = \xi_{it}(y^*_{it} - y_{it-1}),$$ \hspace{1cm} (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 5. That is, the special model we consider in this section corresponds to that in Calvo (1983) with:

\[
\begin{align*}
Pr(\xi_{it} = 0) &= \rho, \\
Pr(\xi_{it} = 1) &= 1 - \rho.
\end{align*}
\]

(3)

It follows from (3) that the expected value of \(\xi_{it}\) is \(1 - \rho\). When \(\xi_{it}\) is zero, the agent experiences inaction; when its value is one, the unit adjusts so as to eliminate the accumulated imbalance. We assume that \(\xi_{it}\) 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:

\[
E[\Delta y_{it} \mid y_{it}^*, y_{it-1}] = (1 - \rho)(y_{it}^* - y_{it-1}),
\]

(4)

so that \(\rho\) represents the degree of inertia of \(\Delta y_{it}\). When \(\rho\) 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 \(\rho\) 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 \(w_i, i = 1, 2, \ldots, n\), 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 \(t\), \(y_{t}^N\), as:

\[
y_{t}^N \equiv \sum_{i=1}^{n} w_i y_{it}.
\]

Similarly we define the value of the aggregate reset value, \(y_{t}^{N*}\), as

\[
y_{t}^{N*} \equiv \sum_{i=1}^{n} w_i y_{it}^*.
\]

Technical Assumptions (Shocks)

Let \(\Delta y_{it}^* \equiv v_{it}^A + v_{it}^I\), where the absence of a subindex \(i\) denotes an element common to all units. We assume:

1. The \(v_{it}^A\)s are i.i.d. with mean \(\mu_A\) and variance \(\sigma_A^2 > 0\).
2. The \(v_{it}^I\)s are independent (across units, over time, and with respect to the \(v^A\)s), identically distributed with zero mean and variance \(\sigma_I^2 > 0\).
3. The $\xi_{it}$’s are independent (across units, over time, and with respect to the $v^A_t$’s and $v^I_t$’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 $\Delta y^\infty_t$ becomes:

$$\Delta y^\infty_t = (1 - \rho)(y^\infty_{t-1} - y^\infty_{t-1}).$$

(5)

Taking first differences yields

$$\Delta y^\infty_t = \rho \Delta y^\infty_{t-1} + (1 - \rho)\Delta y^\infty_{t-1},$$

(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 $v^A_t$ and $v^I_{it}$’s, so that the change in the aggregate reset value of $y$, $\Delta y^\infty_{t}$, is generated by:

$$b(L)\Delta y^\infty_{t} = \epsilon_t,$$

where the $\epsilon_t$’s are i.i.d and $b(L) \equiv 1 - \sum_{i=1}^{q} b_i L^i$ defines a stationary AR(q) for $\Delta y^\infty_{t}$. Assuming Technical Assumption 3 holds we have

$$\Delta y^\infty_t = \rho \Delta y^\infty_{t-1} + (1 - \rho)\Delta y^\infty_{t-1},$$

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

$$(1 - \rho L)b(L)\Delta y^\infty_t = (1 - \rho)\epsilon_t.$$  

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 $\rho$, 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 trend of using quadratic loss functions in economics was initiated by Holt et al. (1961) and continued by Tinsley (1971), Sims (1974) and Sargent (1978).
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.

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 \( \rho \), 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 (\( \rho, \mu_A, \sigma_A \) and \( \sigma_I \)) and \( N \).

**Proposition 1 (Aggregate Bias)**

\[
\hat{\rho} \text{ denote the OLS estimator of } \rho \text{ in } \\
\Delta y^N_t = \text{const.} + \rho \Delta y^N_{t-1} + e_t. \quad (7)
\]

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

\[
\text{plim}_{T \to \infty} \hat{\rho}^N = \frac{K}{1 + K} \rho, \quad (8)
\]

with

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

It follows that:

\[
\lim_{N \to \infty} \text{plim}_{T \to \infty} \hat{\rho}^N = \rho. \quad (10)
\]

**Proof** See Appendix E.  

Statement (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 gap between desired and actual $y$, and a larger amount of inertia as captured by the fraction of agents that do not adjust in any given period, $\rho$.

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, $N$.

### Table 1: Slow Convergence

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>Frequency</th>
<th>Effective number of agents ($N$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>400</td>
</tr>
<tr>
<td>Prices</td>
<td>monthly</td>
<td>0.257</td>
</tr>
<tr>
<td>Employment</td>
<td>quarterly</td>
<td>0.373</td>
</tr>
<tr>
<td>Investment</td>
<td>annual</td>
<td>0.179</td>
</tr>
<tr>
<td>Prices</td>
<td>monthly</td>
<td>0.072</td>
</tr>
<tr>
<td>Employment</td>
<td>quarterly</td>
<td>0.184</td>
</tr>
<tr>
<td>Investment</td>
<td>annual</td>
<td>0.021</td>
</tr>
</tbody>
</table>

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 $\hat{\rho}_\infty \equiv \text{plim}_{T \to \infty} \hat{\rho}$ obtained from Proposition 1. The fourth-sixth rows shows 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: $\hat{\rho} = 0.86$, $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$. Parameters for employment: $\hat{\rho} = 0.60$, $\mu_A = 0.005$, $\sigma_A = 0.03$, $\sigma_I = 0.25$. Parameters for investment: $\hat{\rho} = 0.85$, $\mu_A = 0.12$, $\sigma_A = 0.056$, $\sigma_I = 0.50$. 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 $\hat{\rho} = 0.86$, in line with the median frequency of price adjustments for regular prices reported in Klenow and Kryvtsov (2008).\(^5\) Values for $\mu_A$ and $\sigma_A$ are taken from Bils and Klenow (2004), while $\sigma_I$ consistent with the value estimated in Caballero et al (1997).\(^6\) The table shows that the bias remains significant even for

---

\(^4\) See Appendix D for the definition and main properties of the expected response time.

\(^5\) The average over the eight median frequencies reported by Nakamura and Steinsson for regular price changes suggest taking $\hat{\rho} = 0.89$ which leads to a somewhat larger bias.

\(^6\) To go from the $\sigma_I$ 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
N = 10,000, which corresponds, approximately, to the effective number of prices used to calculate the CPI. In this case, the main reason for the bias is the high value of \( \sigma_I/\sigma_A \).

The second row in 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 \( \mu_A, \sigma_A, \sigma_I \) and \( \rho \). 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 \( \rho, \mu_A \) and \( \sigma_A \), are from Caballero, Engel, and Haltiwanger (1995), and \( \sigma_I \) is consistent with that found in Caballero et al. (1997). 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 \( \rho \), a high \( \mu_A \) (mostly due to depreciation) and a large \( \sigma_I \) (relative to \( \sigma_A \)).

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

\[
p^*_I(t) = (w_t - a_{IT}) + (1 - \alpha_L)l^*_I(t)
\]

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 \( \alpha_L \) is the labor share. It is straightforward to see that as long as the main source of idiosyncratic variance is demand, which we assume, \( \sigma_{l^*_I} \approx (1 - \alpha_L)\sigma_{l^*_T} \).

\( \sigma_I \) is a parameter that is not directly observable in practice. However, one can use a large number of observations to estimate it. In particular, if \( \sigma_I \) is estimated from employment data, it is not clear how to use it to infer \( \sigma_I \) from capital data. A more plausible way to estimate \( \sigma_I \) is to use a large number of observations to estimate it from capital data. A more plausible way to estimate \( \sigma_I \) is to use a large number of observations to estimate it from capital data. A more plausible way to estimate \( \sigma_I \) is to use a large number of observations to estimate it from capital data.

---

7The 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).

8To go from the \( \sigma_I \) 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 \( \sigma_{l^*_K} \approx \sigma_{l^*_T} \).
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 $\rho$,

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

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 $\rho$. 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 $\rho$. To see why this is so, consider the covariance of $\Delta y_t$ and $\Delta y_{t-1}$, noting that, because adjustment is complete whenever it occurs, we may re-write (2) as:

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

where $l_t$ denotes the number of periods since the last adjustment took place, (as of period $t$). 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.

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

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

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:

$$I_k \equiv E_t \left[ \frac{\partial \Delta y_{t+k}}{\partial \Delta y^*_t} \right].$$
It follows from Proposition 1 that the impulse response of $\Delta y$ to $\Delta y^*$ inferred from a linear time-series model estimated for an individual series of $\Delta y$ will be equal to one upon impact and zero for higher lags.

To calculate the correct impulse response, we note that $\Delta y_{t+k}$ responds to $\Delta y^*_t$ 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).9

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 $E[\Delta y_t \Delta y_{t+1}] = 0$ and therefore $\Delta y_{t+1} - E[\Delta y_{t+1} | \Delta y_t] = \Delta y_{t+1}$ so that the Wold innovation at time $t+1$, $\Delta y_{t+1}$, differs from the innovation of economic interest, $\Delta y^*_t$.

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

---

9 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:

\[
\text{plim}_{T \to \infty} \hat{\rho} = \frac{\text{Cov}(\Delta y_{Nt}, \Delta y_{Nt})}{\text{Var}(\Delta y_{Nt})} = \frac{\sum_i w_i^2 \text{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) + \sum_{i \neq j} w_i w_j \text{Cov}(\Delta y_{it}, \Delta y_{2,t-1})}{\sum_i w_i^2 \text{Var}(\Delta y_{it}) + \sum_{i \neq j} w_i w_j \text{Cov}(\Delta y_{it}, \Delta y_{2,t})},
\]

and since \( N = \frac{1}{\sum_i w_i^2} \) and \( \sum_i w_i = 1 \):

\[
\text{plim}_{T \to \infty} \hat{\rho} = \frac{NCov(\Delta y_{it}, \Delta y_{i,t-1}) + N(N-1)Cov(\Delta y_{it}, \Delta y_{j,t-1})}{N\text{Var}(\Delta y_{it}) + N(N-1)Cov(\Delta y_{it}, \Delta y_{j,t})},
\]

(13)

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

<table>
<thead>
<tr>
<th></th>
<th>Cov((\Delta y_{it}, \Delta y_{i,t-1}))</th>
<th>Cov((\Delta y_{it}, \Delta y_{j,t-1}))</th>
<th>Var((\Delta y_{it}))</th>
<th>Cov((\Delta y_{it}, \Delta y_{j,t}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumpy ((\mu_A = 0)):</td>
<td>0</td>
<td>(\frac{1-\rho}{1+\rho} \sigma_A^2)</td>
<td>(\sigma_i^2 + \sigma_j^2)</td>
<td>(\frac{1-\rho}{1+\rho} \sigma_A^2)</td>
</tr>
<tr>
<td>Lumpy ((\mu_A \neq 0)):</td>
<td>(-\rho \mu_A^2)</td>
<td>(\frac{1-\rho}{1+\rho} \sigma_A^2)</td>
<td>(\sigma_i^2 + \sigma_j^2 + \frac{2\rho}{1-\rho} \mu_A^2)</td>
<td>(\frac{1-\rho}{1+\rho} \sigma_A^2)</td>
</tr>
</tbody>
</table>

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, \( \hat{\rho} \) 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:

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

\[
\text{Cov}(\Delta y_{it}, \Delta y_{j,t}) = \frac{1-\rho}{1+\rho} \sigma_A^2,
\]

and we see that the ratio of the two between-agents covariance terms is indeed \( \rho \). 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,

\[10\] 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
\[ \text{Var}(\Delta y_{1,t}) = \sigma_A^2 + \sigma_I^2, \]
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 \( \rho \) 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 \( \Delta y_t \) and \( \Delta y_{t-1} \) 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 \( \Delta y_t \) is equal to \( \mu_A \), 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 \( \mu_A \). It follows that a value of \( \Delta y_t \) above average suggests that it is likely that the agent did not adjust in \( t-1 \), implying that \( \Delta y_{t-1} \) is likely to be smaller than average. Similarly, a value of \( \Delta y_t \) below average suggests that it is likely that the agent adjusted in period \( t-1 \), and \( \Delta y_{t-1} \) 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 \( \mu_A \) times the number of periods).

Summing up, linear time-series models use a combination of self- and cross-covariance terms to estimate the microeconomic speed of adjustment. Inaction biases the self-covariance terms toward infinitely fast adjustment (or even further when \( \mu_A \neq 0 \)). It follows that the ability to recover the true value on \( \rho \) will depend on the cross-covariance terms playing a dominant role. Yet these terms recover \( \rho \) thanks to the common components in the adjustment of different units in consecutive periods, thus their contribution when estimating \( \rho \) will be smaller when adjustment is less frequent (larger \( \rho \)), 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 target \( y^* \). Two alternative approaches—one based on an ARMA representation of \( \Delta y_t^N \), the other on instrumental variables—are discussed in Appendix A.

So far we have assumed that the sluggishness parameter \( \rho \) is estimated using only information
on the economic series of interest, \( y \). Yet often the econometrician can resort to a proxy for the target \( y^* \). 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 \( \Delta y_t^* \).

Equation (14) hints at a procedure for correcting the bias. Since the regressors are orthogonal, from Proposition 1 we have that the coefficient on \( \Delta y_{t-1} \) will be biased downward. By contrast, the true speed of adjustment can be estimated directly from the parameter estimate associated with \( \Delta y_t^* \), 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 \( \rho \) 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 \( \Delta y_t^{*N} \) denotes the average shock in period \( t \), \( \sum w_i \Delta y_i^{*t} \). Then, if (15) is estimated via OLS, and \( K \) defined in (9),

(i) without any restrictions on \( b_0 \) and \( b_1 \):

\[
\lim_{T \to \infty} \hat{b}_0 = \frac{K}{1 + K} \rho, \tag{16}
\]

\[
\lim_{T \to \infty} \hat{b}_1 = 1 - \rho; \tag{17}
\]

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

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

**Proof** See Appendix C.

Proposition 2 entails the general message that constructing a proxy for the target 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 de-
partures from the assumptions we have made so far: $y^*$ does not follow a random walk, agents’ decisions are strategic complements, the probability of adjusting is state-dependent and agents’ adjustment decisions are lumpy but spread out over time (‘time-to-build’). We show that this bias continues be present in all of these cases.

4 Applications

So far we have established the existence of the missing persistence bias theoretically and have 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. For both applications we provide evidence of the presence of the bias and then correct for it using the approach from Section 4.1. 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. 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, 2011). 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.

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, $\lambda$s, estimated from retail level pricing data, and the coefficient $\lambda$.
\( \rho_s \) estimated via OLS from the following regression using the sectoral inflation series \( \pi_{st} \):

\[
\pi_{st} = \rho_s \pi_{s,t-1} + \epsilon_{st}. \tag{18}
\]

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 \( \rho_s \) and that once we correct for this bias the systematic difference between \( \hat{\rho}_s \) and \( 1 - \hat{\lambda}_s \) disappears.

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 propose a new methodology to estimate sectoral shocks, using the CPI micro database, based on repeat price changes. The methodology is analogous to the repeat purchase methodology used to calculate housing price indices. Finally, we use this methodology to estimate sectoral pricing shocks from the CPI database and then use the bias correction approach from Section 3.4 to obtain estimates for \( \rho_s \) that are immune to the missing persistence bias. We find that the bias correction method does a good job, that is, we find that \( \hat{\rho}_s \approx 1 - \hat{\lambda}_s \).

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.\(^{13}\) The number of sectors we consider is 66.

The calibration we use is standard and the details are relegated to Appendix G. 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 from our simulations are summarized in Figure 1. The prediction of the Calvo model is shown by

\(^{13}\)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.
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.

[I WOULD REMOVE THESE METHODS THAT DIDNT WORK FROM THE CORE OF THE PAPER]

Having established that the missing persistence bias is potentially relevant, next we study the extent to which the bias correction methods from Section 3.4 reduce the bias. In simulations, the ARMA correction and the IV approach were too fragile, thus we decided to use micro data to estimate the sectoral shocks and apply the correction procedure described in Section 3.4.

As a building block toward the measure we will actually use, we first consider a particular case where we estimate the sectoral shock, in sector \( s \) in period \( t \), \( v_{st} \), as the average price change of price-setters in sector \( s \) at time \( t \) that adjusted in both periods \( t - 1 \) and \( t \). Under the assumptions of the Calvo model considered in Section 3, this estimator is equal to the sum of the sectoral shock and the average of firm specific (idiosyncratic) shocks across the pricesetters that are considered. The estimator is therefore unbiased and its standard deviation is equal to \( \sigma_I / \sqrt{\tilde{n}_{st}} \) with \( \tilde{n}_{st} \) equal to the number of firms that adjusted both in \( t - 1 \) and \( t \).

A practical drawback of this intuitive approach is that only establishment level price changes corresponding to periods where there also was a price adjustment in the previous period are used, that is, on average we will only have \( n_{st} \lambda_s^2 \) observations to estimate \( v_{st} \), where \( n_{st} \) denotes the average (over time) number of price-setters in sector \( s \) at time \( t \) and \( \lambda_s \) the (average) fraction that adjusted...
prices in any given period. This means that there can be many periods with no (or very few) observations to calculate an estimate for \( v_{st} \), especially in sectors with a low frequency of adjustment. In the CPI research database, the median number of observations and frequency adjustment across sectors is 132 and 0.079, respectively so in practice this limitation is severe. Thus we propose next a modified approach that uses all price changes.

Concretely, if firm \( f \) in sector \( s \) adjusted in periods \( t \) after last adjusting in period \( t - k \) then its price change can be decomposed into the following sum:

\[
\pi_{fst} = v_{st} + v_{s,t-1} + ... + v_{s,t-k+1} + e_{fst}
\]  (19)

where \( v_{s,t-j} \) denotes the common (across price-setters in sector \( s \)) shock in period \( t \) and \( e_{fst} \) denotes the sum of \( k \) idiosyncratic shocks hitting the firm between \( t - k + 1 \) and \( t \).

We describe next how to estimate sectoral shocks in sector \( s \) when we observe price changes during periods 1 through \( T \): \( v_{s,1}, v_{s,2}, \ldots, v_{s,T} \); based on the linear system obtained from considering (19) for all price adjustments in sector \( s \) during the \( T \) periods of interest. The method is analogous to repeat purchase methods used to derive housing price indices such as the well known Case-Shiller index.

Observed price changes in (19) are the dependent variable, sectoral shocks are the parameters that are estimated and the row of the design matrix that corresponds to \( \pi_{fst} \) in (19) is a row with ones in positions \( t - k + 1 \) through \( t \) and zeros elsewhere. Though similar in spirit to the simpler method we described above, a virtue of this approach is the number of observations available for estimation is much larger, \( n_s \lambda_s \) rather than \( n_s \lambda_s^2 \).

Under the Technical Assumptions from Section 2, the error term satisfies standard orthogonality restrictions. Also, the variance of the error term is \( k \sigma_j^2 \) with \( k \) equal to the number of periods since the last price adjustment for the price change observation under consideration. Because the researcher observes \( k \), it follows that weighted least squares provides efficient (and unbiased) estimates for the sectoral shocks.

Our “repeat price change” methodology for estimating shocks is conceptually similar to Bils, Klenow and Malin’s (2012) “reset price inflation” approach. Both methodologies make inferences about the underlying shocks using information contained in observed price changes. The above digression shows that our measure is more efficient under the Calvo model assumptions, simulation results show that both measures have similar performance in \( Ss \)-type settings.

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

\[
\pi_{st} = \beta_s \pi_{s,t-1} + \gamma_s v_{st} + e_{st}
\]  (20)

Proposition 2 implies that if we estimate \( \beta \) and \( \gamma \) in the above equation without imposing any constraints across them, \( \hat{\gamma}_s \) will be an unbiased estimate of actual fraction of adjusters \( \lambda_s \).
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 red circles in Figure 1, where each circle represents one corrected estimate for \( 1 - \lambda_s \) based on estimating a linear time series model for sectoral inflation data. All predictions now lie close to the Calvo prediction (the solid line).\(^{14}\)

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. 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.\(^ {15}\) The database contains thousands of individual “quote-lines” with price observations for many months. In our data set, an average month contains about 10,000 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 a 16 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. (Nevertheless, we have also repeated the analysis including product substitutions and found similar results.)

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 \( \lambda_s^{\text{VAR}} = 1 - \hat{\beta}_s \). As in Bils and Klenow (2004), we find that \( \hat{\beta}_s \ll 1 - \lambda_s^{\text{micro}} \), where \( \lambda_s^{\text{micro}} \) denotes the true frequency of adjustment, estimated from the micro level quote-lines. Next we estimate equation (20) 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 \( \lambda_s^c \), where the superindex \( c \) stands for “corrected”. To gauge the extent to which the \( \lambda_s^c \) 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:

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

\(^{14}\)The fact that they lie a little bit above the line may be due to the presence of strategic complementarities.

\(^{15}\)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 NY, LA and Chicago, but these items are not representative, so we restrict our monthly analysis to these three cities.
Here $\eta$ 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 $\alpha$ will be large while $\eta$ won’t be significantly different from zero. By contrast, if the bias reduction is proportional to the actual bias, we expect an estimate of $\eta$ that is significantly positive, taking values close to one if the bias completely disappears.

Table 4: Bias-Correction Estimation

<table>
<thead>
<tr>
<th></th>
<th>Multi-sector Calvo Model (simulations)</th>
<th>CPI database (actual data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta$</td>
<td>1.038***</td>
<td>1.059***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.003</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Observations</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.99</td>
<td>0.93</td>
</tr>
</tbody>
</table>

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 $\eta$ is not statistically different from one. This suggests that the departure from the Calvo model found in Figure 2 in BK is probably driven by the missing persistence bias.

This example shows the bias is relevant at the sectoral level and that an innovative use of micro data 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. In a recent paper, 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.

To understand BGM’s approach, we must first introduce some terminology. Define $\Pi_t$ 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. They assume that $\Pi_t$ can be decomposed into the sum of small number $K$ of common factors, $C_t$, and a sectoral component, $e_t$:

$$\Pi_t = \Lambda C_t + e_t$$ (21)
Here $\Lambda$ is a matrix of factor loadings for the common factors. These factor loadings are allowed to differ across sectors, so that $\Lambda$, $C_t$ and $e_t$ are $SxK$, $Kx1$ and $Sx1$ matrices.

BGM extract $K$ principal components from the large data set $\Pi_t$ to obtain consistent estimates of the common factors.\(^{16}\) Next they regress each sectoral inflation series on the common factors, denoting the predicted aggregate component by $\pi_{st}^{\text{agg}}$, and the residual, which captures the sector specific component, by $\pi_{st}^{\text{sect}}$:

$$\pi_{it} = \Lambda_i' C_t + e_{it} \quad (22)$$

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 have a diffuse effect on all data series—from those due to sector-specific conditions represented by the term $e_{it}$. To calculate IRFs with respect to the common and sectoral shocks, BGM fit separate AR(13) processes to $\pi_{st}^{\text{agg}}$ and $\pi_{st}^{\text{sect}}$ series and measure the persistence of shocks by the sum of the 13 AR coefficients.\(^{17}\)

To start, we reproduce their benchmark results using our 66 series. The results are shown in Table 5. Both the mean and median persistence of the aggregate component is much larger than the mean and median persistence of the sectoral component. This suggests that sectoral inflation rates respond much faster to sectoral shocks than to aggregate shocks in the U.S. A subsequent paper by Mackoviak, Moench and Wiederholt (2011), using a different methodology, found similar results using the CPI data. Both papers conclude that this difference in persistence is strong evidence in favor of sticky-information models.

<table>
<thead>
<tr>
<th>Persistence measure</th>
<th>$\pi_{st}^{\text{agg}}$</th>
<th>$\pi_{st}^{\text{sect}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.92</td>
<td>-0.07</td>
</tr>
<tr>
<td>median</td>
<td>0.99</td>
<td>-0.01</td>
</tr>
<tr>
<td>std</td>
<td>0.08</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Notice, however, that BGM measure persistence of each component by regressing each component on lags of itself. Since the underlying prices adjust infrequently and there are not many prices underlying these sectoral inflation series, could BGM’s results be driven by the missing persistence bias? We explore this possibility in the remainder of this section.

To investigate this hypothesis, we use the same shock measures that we computed from CPI micro data that were discussed in depth in Section 6.1. That is, we have data for 66 sectoral inflation series from the CPI for the period 1988:03-2007:12.

\(^{16}\)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.

\(^{17}\)For example, under the Technical Assumptions this sum equals to $\rho$. 
Define $V_t$ as the $S \times 1$ 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 $m^k_t$, $k = 1, 2, \ldots, K$. 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:

$$v_{st} = \sum_{k=1}^{K} \sum_{j \geq 0} \gamma_{sj}^k m^k_{t-j} + x_{st}.$$ 

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

So far we have $K$ aggregate shock components, $m^k_t$, and a sectoral shock, $x_{st}$, for each of 66 sectors from the CPI. 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:

$$\pi_{st} = \sum_{k=1}^{K} \sum_{j \geq 0} \eta_{sj}^k m^k_{t-j} + \sum_{j \geq 0} \nu_{sj} x_{s,t-j}. \quad (23)$$

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 to each of the $K$ aggregate shocks and summarize the $K$ response times to aggregate shocks by their median:

$$\tau_{s,ag}^k = \frac{\sum_{j \geq 0} \eta_{sj}^k} {\sum_{j \geq 0} \eta_{sj}^k},$$

$$\tau_{s,sec} = \frac{\sum_{j \geq 0} \nu_{sj}^k} {\sum_{j \geq 0} \nu_{sj}^k},$$

$$\tau_{s} = \text{median}_k \tau_{s,k}.$$ 

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

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 four combinations for the number of principal components and distributed lags.

Columns (1) and (2) show the estimated response times when using the BGM approach. In all specifications the persistence of the aggregate shocks is significantly larger than the persistence of the sectoral shock, just as BGM asserted. In three out of four specifications the estimated response time to aggregate shocks is close to twice as large.

Columns (3) and (4) show the results from using the approach outlined above to correct for the missing persistence bias. Our results are robust to ignoring distributed lags of common components yet we believe it is more realistic to include these components.
Table 6: The response of sectoral inflation rates to aggregate and idiosyncratic shocks

<table>
<thead>
<tr>
<th>PCs</th>
<th>nlags</th>
<th>BGM agg</th>
<th>BGM sec</th>
<th>bias corrected agg</th>
<th>bias corrected sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>1.79</td>
<td>0.87</td>
<td>5.58</td>
<td>5.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.32)</td>
<td>(0.24)</td>
<td>(0.36)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1.37</td>
<td>1.10</td>
<td>5.91</td>
<td>5.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.45)</td>
<td>(0.25)</td>
<td>(0.38)</td>
<td>(0.60)</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>1.70</td>
<td>0.96</td>
<td>5.99</td>
<td>5.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(0.26)</td>
<td>(0.34)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>1.50</td>
<td>0.76</td>
<td>5.92</td>
<td>6.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.34)</td>
<td>(0.23)</td>
<td>(0.26)</td>
<td>(0.50)</td>
</tr>
</tbody>
</table>

missing persistence bias. The average response time, across the four specifications we consider, increases on average from 0.92 to 5.76 months for sectoral shocks and from 1.59 to 5.85 months for aggregate shocks. That is, response times increase substantially once we correct for the missing persistence bias, by a factor of more than 6 for sectoral shocks and almost 4 for aggregate shocks. This increase provides evidence in favor of the empirical relevance of the missing persistence bias, which we would expect to be present for both response times.

The results in columns (3) and (4) also show that after correcting for the missing persistence bias, 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 less than 1/10 of a month.

5 Conclusion

[Coming soon]

\[19\] Columns (3) and (4) infer the response times from estimating (23). As a robustness check we estimated \( K \) regressions for each sector, including one component of the aggregate shock at a time in each regression, instead of including all components simultaneously. This approach leads to similar results.
References


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 \(\Delta y^N_t\).

**Proposition 3 (ARMA Representation)**

Consider the assumptions and notation of Proposition 1. We then have that \(\Delta y^N_t\) follows the following ARMA(1,1) process:

\[
\Delta y^N_t = \rho \Delta y^N_{t-1} + (1 - \rho)[\epsilon_t - \theta \epsilon_{t-1}],
\]

(24)

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

**Proof** See Appendix C.

Using (24) to write \(\Delta y^N_t\) as an infinite moving average shows that its impulse response to \(\epsilon\)-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 \(v^A_t\), because \(\epsilon_t\) in (24) is not \(v^A_t\). 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, \(\rho\): Estimate an ARMA(1,1) process (24) and read off the estimate of \(\rho\) (and the true impulse response) from the estimated AR-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 \(\Delta y^N_t\) is close to an i.i.d. process which means that \(\theta\) and \(\rho\) 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, \(\rho\).

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).

\[\text{Scaling the right hand side term by } (1 - \rho) \text{ is inoccuous but useful in what follows.}\]
A.2 Instrumental Variables

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

$$\Delta y^*_t = \text{const.} + \rho \Delta y^*_t + e_t.$$  

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

**Proposition 4 (Instrumental Variables)**

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

$$\Delta y^*_t = \text{const.} + \rho \Delta y^*_{t-1} + e_t.$$  

*By contrast, $\Delta y^*_{t-1}$ 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 $\Delta y^*$) are independent across periods. Thus the correlation between $\Delta y_t$ and $\Delta y_{t-1}$ 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 $S_s$-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.\(^{21}\)

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 $\rho$ 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.

\(^{21}\)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 Relating the i.i.d. Assumption

In Section 3 we have assumed that \( \Delta y^* \) 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 \( \Delta y^*, \nu_t^A \) and \( \nu_t^I \), follow AR(1) processes with the same first-order autocorrelation \( \phi \). The case we considered in the main text corresponds to \( \phi = 0 \). We show in Appendix E that, with a continuum of agents, \( \Delta y_t^{\infty} \) 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 \( \varepsilon_t \) proportional to \( \nu_t^A \) and \( \beta \) denoting the agent’s discount factor.\(^{22}\)

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 \( \Delta y^* \) is not i.i.d. but also knows the exact value of the first order autocorrelation, \( \phi \), as well as \( \beta \), and estimates \( \rho \) via maximum likelihood from

\[
(\Delta y_t^N - \phi \Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi \Delta y_{t-2}^N) + \varepsilon_t - \beta \rho \phi \varepsilon_{t-1}.
\]

The only source of bias is that the researcher ignores the fact that because the actual aggregate is not i.i.d. but also knows the exact value of the first order autocorrelation, \( \phi \), as well as \( \beta \), and estimates \( \rho \) via maximum likelihood from

\[
(\Delta y_t^N - \phi \Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi \Delta y_{t-2}^N) + \varepsilon_t - \beta \rho \phi \varepsilon_{t-1}.
\]

It follows from Appendix E that, with a continuum of agents, \( \Delta y_t^{\infty} \) follows the following stationary ARMA(2,1) process:

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

with \( \varepsilon_t \) proportional to \( \nu_t^A \).

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.\(^{24}\)

\(^{22}\)With the notation of Section 2 we have \( b(L) = (1 - \phi L)/(1 - \beta \rho \phi L) \).

\(^{23}\)Simulations show that the bias disappears if we estimate \( \Delta y_t^N - \phi \Delta y_{t-1}^N = \text{const.} + \rho(\Delta y_{t-1}^N - \phi \Delta y_{t-2}^N) + \varepsilon_t - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} \) with no constraints on \( \gamma_1 \) and \( \gamma_2 \). This suggests that the random walk assumption can be relaxed in Proposition 3.

We thank Juan Daniel Díaz for this insight.

\(^{24}\)Nonetheless, results are qualitatively similar if we work with prices.
Table 7: SLOW CONVERGENCE

Estimated Half-Life and Expected Response Time $\Delta y^*$ follows an AR(1)

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>100</th>
<th>400</th>
<th>1,000</th>
<th>4,000</th>
<th>10,000</th>
<th>40,000</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.252</td>
<td>0.466</td>
<td>0.769</td>
<td>1.724</td>
<td>2.639</td>
<td>3.794</td>
<td>4.596</td>
</tr>
<tr>
<td>0.1</td>
<td>0.246</td>
<td>0.440</td>
<td>0.723</td>
<td>1.683</td>
<td>2.659</td>
<td>3.841</td>
<td>4.615</td>
</tr>
<tr>
<td>0.2</td>
<td>0.296</td>
<td>0.426</td>
<td>0.686</td>
<td>1.671</td>
<td>2.646</td>
<td>3.852</td>
<td>4.644</td>
</tr>
<tr>
<td>0.3</td>
<td>0.379</td>
<td>0.459</td>
<td>0.661</td>
<td>1.615</td>
<td>2.651</td>
<td>3.882</td>
<td>4.690</td>
</tr>
<tr>
<td>0.4</td>
<td>0.529</td>
<td>0.564</td>
<td>0.662</td>
<td>1.589</td>
<td>2.697</td>
<td>3.993</td>
<td>4.764</td>
</tr>
<tr>
<td>0.5</td>
<td>0.751</td>
<td>0.767</td>
<td>0.801</td>
<td>1.416</td>
<td>2.704</td>
<td>4.064</td>
<td>4.887</td>
</tr>
</tbody>
</table>

First six rows report the average estimate of the half-life of a shock. The parameter $\rho$ is estimated via maximum likelihood from $(\Delta y^N_t - \phi \Delta y^N_{t-1}) = \text{const.} + \rho (\Delta y^N_{t-1} - \phi \Delta y^N_{t-2}) + e_t - \beta \rho e_{t-1}$ with $\beta$ and $\phi$ 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 $\Delta y^N_t = \sum_{k=0}^{\infty} \psi_k y_{t-k}$ is the (infinite) MA representation of $\Delta y^N_t$ 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 $(\phi + \rho - 2\phi \rho)/(1 - \phi - \rho) - \beta \rho \phi/(1 - \beta \rho \phi)$ (see Appendix D). Parameters (monthly pricing data): $\rho = 0.86$, $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$, $\beta = 0.96^{1/12}$.

Table 8: SLOW CONVERGENCE

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

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>100</th>
<th>400</th>
<th>1,000</th>
<th>4,000</th>
<th>10,000</th>
<th>40,000</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.493</td>
<td>0.374</td>
<td>0.287</td>
<td>0.198</td>
<td>0.172</td>
<td>0.158</td>
<td>0.150</td>
</tr>
<tr>
<td>0.7</td>
<td>0.599</td>
<td>0.448</td>
<td>0.328</td>
<td>0.210</td>
<td>0.177</td>
<td>0.158</td>
<td>0.150</td>
</tr>
<tr>
<td>0.8</td>
<td>0.712</td>
<td>0.533</td>
<td>0.385</td>
<td>0.231</td>
<td>0.186</td>
<td>0.161</td>
<td>0.150</td>
</tr>
<tr>
<td>0.9</td>
<td>0.843</td>
<td>0.646</td>
<td>0.469</td>
<td>0.269</td>
<td>0.205</td>
<td>0.169</td>
<td>0.150</td>
</tr>
<tr>
<td>1.0</td>
<td>0.982</td>
<td>0.856</td>
<td>0.697</td>
<td>0.410</td>
<td>0.279</td>
<td>0.188</td>
<td>0.150</td>
</tr>
</tbody>
</table>

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

28
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 \( \Delta y \) become infinite.\(^{25}\) Reported estimates assume the researcher knows the value of \( \phi \) in the AR(1) process but believes \( N = \infty \), and therefore estimates \( \rho \) 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 \( \phi \) 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 \( \phi \) 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 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 \( \epsilon_t \). Aggregate inflation, \( \pi_t \), then follows an AR(1) process

\[
\pi_t = \phi \pi_{t-1} + (1-\phi)\epsilon_t
\]

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:

\[
\Delta \log p^*_t = \phi \Delta \log p^*_{t-1} + c(\epsilon_t - \rho \epsilon_{t-1}),
\]

with \( c = (1 - \phi)/(1 - \rho). \(^{26}\)

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.

\(^{25}\) Also, if we report the half-life and expected response time for \( y \) instead of \( \Delta y \), 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 \( \phi \) tends to one.

\(^{26}\) In the notation of Section 2 we have \( b(L) = (1 - \phi L)/(1 - \rho L) \).
Table 9: SLOW CONVERGENCE AND STRATEGIC COMPLEMENTARITIES

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \phi )</th>
<th>Effective number of agents (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>0.8600</td>
<td>0.8600</td>
<td>0.257</td>
</tr>
<tr>
<td>0.8600</td>
<td>0.9442</td>
<td>0.268</td>
</tr>
<tr>
<td>0.8600</td>
<td>0.8600</td>
<td>0.072</td>
</tr>
<tr>
<td>0.8600</td>
<td>0.9442</td>
<td>0.081</td>
</tr>
</tbody>
</table>

First two rows show the estimated half-life. The half-life is calculated from \(- \log_2 / \log \hat{\rho}_\infty\) with \(\hat{\rho}_\infty \approx \text{plim}_{T \to \infty} \hat{\rho}\) 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, \(\Delta y_t\). This adjustment takes place only gradually, for example, because of a time-to-build component. We capture this pattern with the process:

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

Now there are two sources of sluggishness in the transmission of shocks, \(\Delta y^*_t\), to the observed variable, \(\Delta y_t\). 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:

\[
\Delta \tilde{y}_t = \rho \Delta \tilde{y}_{t-1} + \nu_t.
\] (27)
Replacing (27) into (26), yields the following linear equation in terms of the observable, $\Delta y_t$:

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

with

$$a_1 = \phi_1 + \rho, \quad a_k = \phi_k - \rho \phi_{k-1}, \quad k = 2, ..., K,$$

$$a_{K+1} = -\rho \phi_K,$$

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

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

**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:

$$\lim_{T \to \infty} \hat{a}_k = \phi_k, \quad k = 1, ..., K,$$

$$\lim_{T \to \infty} \hat{a}_{K+1} = 0.$$  \hfill (30)

**Proof** See Appendix C.

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

## C Proof of Propositions

**Proof of Proposition 1**

In this appendix we prove Proposition 1. The proof uses an auxiliary variable equal to how much unit $i$ adjusts in period $t$ if it adjusts that period. Denoting this variable by $x_{it}$, we have:

$$x_{it} \equiv y^*_t - y_{i,t-1}.$$

Give the Technical Assumptions, we have that $x_{it}$ equals the unit’s accumulated shocks since it last adjusted.

The following expressions characterize the dynamics of $x_{it}$ as well as relating this variable to changes in the variable of interest:

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

$$\Delta y_{it} = \xi_{it} x_{it}. \quad (32)$$
In what follows, subindices $i$ and $j$ denote *different* units.

We first derive the following unconditional expectations:

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

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

\[
\mathbb{E}x_{i,t+1} = \rho \mathbb{E}x_{it} + \mu_A.
\]

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

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

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

\[
\begin{align*}
\mathbb{E}[x_{i,t+1}x_{j,t+1}] &= \mathbb{E}[(1 - \xi_{it})x_{it} + \Delta y_{i,t+1}^*][(1 - \xi_{jt})x_{jt} + \Delta y_{j,t+1}^*] \\
&= \mathbb{E}[(1 - \xi_{it})x_{it}(1 - \xi_{jt})x_{jt}] + \mathbb{E}[\Delta y_{i,t+1}^* (1 - \xi_{jt})x_{jt}] \\
&\quad + \mathbb{E}[(1 - \xi_{it})x_{it}\Delta y_{j,t+1}^*] + \mathbb{E}[\Delta y_{i,t+1}^* \Delta y_{j,t+1}^*] \\
&= \rho^2 \mathbb{E}[x_{it}x_{jt}] + 2\frac{\rho}{1 - \rho} \mu_A^2 + (\mu_A^2 + \sigma_A^2),
\end{align*}
\]

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

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

\[
\begin{align*}
\mathbb{E}[x_{i,t+1}^2] &= \mathbb{E}[(1 - \xi_{it})x_{it}^2] + 2\mathbb{E}[1 - \xi_{it})x_{it}\Delta y_{i,t+1}^*] + \mathbb{E}[(\Delta y_{i,t+1}^*)^2] \\
&= \rho \mathbb{E}[x_{it}^2] + 2\frac{\rho}{1 - \rho} \mu_A^2 + (\sigma_A^2 + \sigma_t^2 + \mu_A^2),
\end{align*}
\]

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:

\[
\begin{align*}
\text{Cov}(\Delta y_{i,t+1}, \Delta y_{it}) &= \mathbb{E}[\Delta y_{i,t+1}\Delta y_{it}] - \mu_A^2 = \mathbb{E}[(1 - \xi_{it})x_{it}\Delta y_{i,t+1}^*] - \mu_A^2 \\
&= (1 - \rho)\mathbb{E}[(1 - \xi_{it})x_{it} + \Delta y_{i,t+1}^*][\xi_{it}x_{it}] - \mu_A^2 \\
&= (1 - \rho)[\mathbb{E}[(1 - \xi_{it})x_{it}^2] + (1 - \rho)\mathbb{E}[\Delta y_{i,t+1}^* \xi_{it}x_{it}] - \mu_A^2] \\
&= -\rho \mu_A^2 - \mu_A^2 = -\rho \mu_A^2,
\end{align*}
\]
where in the crucial step we used that \((1 - \xi_{it})\xi_{it}\) always equals zero.

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

\[
\begin{align*}
\text{Cov}(\Delta y_{t+1}, \Delta y_{jt}) &= E[\xi_{i,t+1}x_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)E[x_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 \\
&= (1 - \rho)E[(1 - \xi_{it})x_{it} + \Delta y_{it+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = \rho(1 - \rho)^2E[x_{it}x_{jt}] + (1 - \rho)\mu_A^2 - \mu_A^2 = \frac{1 - \rho}{1 + \rho}A^2.
\end{align*}
\]

\[
\begin{align*}
\text{Cov}(\Delta y_{it}, \Delta y_{jt}) &= E[\xi_{it}x_{it}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)^2E[x_{it}x_{jt}] - \mu_A^2 = \frac{1 - \rho}{1 + \rho}A^2.
\end{align*}
\]

Finally, the variance term is obtained as follows:

\[
\text{Var}(\Delta y_{it}) = E[\xi_{it}^2x_{it}^2] - \mu_A^2 = E[\xi_{it}x_{it}^2] - \mu_A^2 = (1 - \rho)E[x_{it}^2] - \mu_A^2 = \sigma_A^2 + \sigma_T^2 + \frac{2\rho}{1 - \rho}A^2.
\]

**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:

\[
\text{plim}_{T \to \infty} \hat{\theta}_1 = \frac{\text{Cov}(\Delta y_t - \Delta y_{t-1}, \Delta y_{t}^* - \Delta y_{t-1})}{\text{Var}(\Delta y_{t}^* - \Delta y_{t-1})}.
\]

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

\[
\text{Cov}(\Delta y_t^N, \Delta y_{t}^N) = \frac{1}{N}\text{Cov}(\Delta y_{it}, \Delta y_{it}^*) + \left(1 - \frac{1}{N}\right)\text{Cov}(\Delta y_{it}, \Delta y_{jt}).
\]

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

**Proof of Proposition 3**

To prove that \(\Delta y_t^N\) follows an ARMA(1,1) process with autoregressive coefficient \(\rho\), it suffices to show that the process's autocorrelation function, \(\gamma_k\), satisfies:

\[
\gamma_k = \rho \gamma_{k-1}, \quad k \geq 2. \tag{38}
\]

We prove this next and derive the moving average parameter \(\theta\) by finding the unique \(\theta\) 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 \(\theta\) tends to zero as \(N\) tends to infinity is straightforward.

\footnote{Here we are using Theorem 1 in Engel (1984) characterizing ARMA processes in terms of difference equations satisfied by their autocorrelation function.}
We have:

$$E[\Delta y^N_{i+k} \Delta y^N_t] = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{i,t+k} x_{i,t+k} \xi_{j,t} x_{j,t}]$$

$$= (1 - \rho) \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[x_{i,t+k} \xi_{j,t} x_{j,t}]$$

$$= (1 - \rho) \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[(1 - \xi_{i,t+k-1}) x_{i,t+k-1} + \Delta y^N_{i,t+k} \xi_{j,t} x_{j,t}]$$

$$= (1 - \rho) \rho \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[x_{i,t+k-1} \xi_{j,t} x_{j,t}] + (1 - \rho) \mu^2_A \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{i,t} x_{j,t}]$$

$$= \rho \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{i,t+k-1} x_{i,t+k-1} \xi_{j,t} x_{j,t}] + (1 - \rho) \mu^2_A$$

$$= \rho E[\Delta y^N_{i,t+k-1} \Delta y^N_t] + (1 - \rho) \mu^2_A,$$

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

**Proof of Proposition 4**

We have:

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

Similarly

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

Subtracting the latter from the former and rearranging terms yields

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

with

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

The extra term $\epsilon^N_t$ on the r.h.s. of (40) explains why $\Delta y^N_{t-1}$ is not a valid instrument: $\Delta y^N_{t-1}$ is correlated with $\epsilon^N_t$ because both include $\xi_{i,t-1}$ terms. Of course, $\epsilon^N_t$ 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

$$\text{Var}(\epsilon_t) = \frac{2 \rho}{N} \left[ \sigma^2_A + \sigma^2_I + \frac{1 + \rho}{1 - \rho} \mu^2_A \right].$$

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

**Proof of Proposition 5**
The equation we estimate is:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t,$$  \hspace{1cm} (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 $\Delta y_{t-k}$, $k \geq 1$. 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: $\tau$

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

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

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

Define:

$$I_k \equiv E_t \left[ \frac{\partial \Delta y_{t+k}}{\partial \varepsilon_t} \right] \quad \text{and} \quad \tau \equiv \frac{\sum_{k \geq 0} k I_k}{\sum_{k \geq 0} I_k}. \hspace{1cm} (42)$$

Then:

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

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

**Proposition A1 (\(\tau\) for an ARMA Process)** Assume $\Delta y_t$ follows an ARMA($p,q$):

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

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

Define $\tau$ 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:

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

Proposition A2 (τ for a Lumpy Adjustment Process) Consider $\Delta y_t$ in the simple lumpy adjustment model (12) and $\tau$ defined in (??). Then $\tau = \rho / (1 - \rho)^2$.

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:

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

The expression for $\tau$ now follows easily. ■

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 $\rho$) or zero (with probability $1 - \rho$) (Calvo Model). When the agent adjusts, it chooses $y_{it}$ equal to $\tilde{y}_t$ that solves

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

where $\beta$ denotes the agent’s discount factor and $y_t^*$ denotes an exogenous process.29 We then have

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

It follows that, as $N$ tends to infinity, $y_t^\infty$ satisfies:

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

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

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

28More generally, if the number of periods between consecutive adjustments are i.i.d. with mean $m$, then $\tau = m - 1$. What follows is the particular case where interarrival times follow a Geometric distribution.

29This formulation can be extended to incorporate idiosyncratic shocks.
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.\(^{30}\)

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

with

\[
\hat{y}_t = (1 - \delta) \sum_{k=0}^\infty \delta^k E_t y_{t+k}^*.
\]

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: \( \rho' \) plays the role of \( \rho \) and \( \delta \) 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 \( E_t y_{t+k}^* = \psi^k y_t^* \) and \( y_t^\infty \) follows the following AR(2) process:

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

**b)** Consider the case where \( \Delta y_t^* \) follows an AR(1):

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

with \(|\phi| < 1\). We then have that

\[
E_t y_{t+k}^* = \frac{\phi (1 - \phi^k)}{1 - \phi} \Delta y_t^* + y_t^*
\]

and \( \Delta y_t^\infty \) follows the following ARMA(2,1) process:

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

**Proof** Straightforward.

**F** The case where \( y^* \) is i.i.d.

Assume that

\[
y_t^* = y_t^* A + y_t^* I
\]

\(^{30}\)The expression that follows is equivalent to the partial adjustment formulation:

\[
\Delta y_t^\infty = (1 - \rho') (y_t - y_{t-1}^\infty),
\]
with \( y_{it}^A \) i.i.d. with mean \( \mu_A \) and variance \( \sigma_A^2 \) and \( y_{it}^I \) i.i.d. with zero mean and variance \( \sigma_I^2 \). The \( y_{it}^I \) processes are independent across agents and independent from the aggregate shock process \( y_{it}^A \). 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:

\[
y_{it}^\infty = \rho y_{it-1}^\infty + (1 - \rho)(1 - \beta \rho) y_{it}^A.
\] (49)

We show next that the OLS estimator of \( \rho \) in the regression

\[
y_{it}^\infty = \rho y_{it-1}^\infty + e_t
\] (50)

provides a consistent estimator of \( \rho \) 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

\[
\hat{y}_{it} = (1 - \beta \rho) y_{it}^*.
\]

Using the notation we introduced in Appendix C this implies that

\[
y_t^N = \frac{1}{N} \sum_{i=1}^{N} (1 - \xi_{it}) y_{i,t-1} + (1 - \beta \rho) \frac{1}{N} \sum_{i=1}^{N} \xi_{it} y_{it}^*.
\]

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

\[
y_t^N = \rho y_{t-1}^N + \varepsilon_t
\] (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 \( \varepsilon_t \) differs from the error term in (49), it also is uncorrelated with the regressor \( y_{t-1}^N \) which is all we need for \( \hat{\rho} \) estimated via OLS from (51) to be a consistent estimator for \( \rho \).

### G 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 270 periods, which is the number of periods in the underlying data.

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

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of monthly adjustment:</td>
<td>0.068</td>
<td>0.068</td>
</tr>
<tr>
<td>Fraction price changes &gt; 0:</td>
<td>0.669</td>
<td>0.567</td>
</tr>
<tr>
<td>Average size of increases (%)</td>
<td>7.997</td>
<td>8.305</td>
</tr>
<tr>
<td>Average size of decreases (%)</td>
<td>9.073</td>
<td>8.180</td>
</tr>
<tr>
<td>Std of sectoral inflation</td>
<td>0.004</td>
<td>0.005</td>
</tr>
</tbody>
</table>