MISSING AGGREGATE DYNAMICS AND VAR APPROXIMATIONS OF LUMPY ADJUSTMENT MODELS

David Berger
Northwestern and NBER

Ricardo Caballero
MIT and NBER

Eduardo Engel
Universidad de Chile

November 19, 2018[1]

Abstract

When microeconomic adjustment is lumpy, the estimated persistence inferred from VARs is biased downwards. This "missing persistence bias" decreases with the level of aggregation, but it converges slowly. Because of this slow convergence, assuming infinitely many agents in empirical applications often results in biased estimates. We look at the magnitude of this bias in empirical estimates of impulse response functions and simulation based estimators. We find that the "missing persistence bias" is present and meaningful, and we propose a method for estimating the true speed of adjustment. We illustrate our method's effectiveness with simulated data and two real world applications with pricing data.

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

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

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

1 INTRODUCTION

Macroeconomists want to know how aggregate variables respond to shocks. To measure these dynamics, we often employ a vector autoregression (VAR) at some point. The characterization may stop there, or researchers may go on to estimate underlying model parameters. Our paper urges caution. When the microeconomic adjustment process underlying an aggregate variable is lumpy, conventional VAR procedures often underestimate persistence. We refer to this problem as "missing persistence bias" (hereafter simply "bias" or "the bias").

This bias is less problematic with more highly aggregated data. Linear models miss any persistence that might be present in an individual series, while estimates on aggregate data with infinitely many agents are unbiased. We look at the cases in between. Convergence is slow, and in practice the bias is meaningful for sectoral data and, sometimes, aggregate series as well.

We highlight two significant implications of this bias for applied research. First, VAR estimates of impulse response functions often underestimate the persistence of shocks. We can solve this problem by using projection methods when estimating IRFs (as in Jorda, 2005), since these are immune to the bias. Second, it is important to simulate the true number of agents when using indirect inference and simulated method of moments to estimate/calibrate macroeconomic models. The common practice of using a very large number of agents is likely to underestimate the persistence of shocks.

We provide two detailed applications where we correct for this bias. In the first application, we explain why estimates for the speed of adjustment of sectoral prices obtained using direct measures are much lower than those obtained with standard linear time-series models, thereby potentially solving a puzzling finding in Bils and Klenow (2004). In this application we can measure the size of the missing persistence bias, and we find that our bias correction procedure works well in practice: linear time series models deliver estimates in line with those obtained from unbiased nonlinear methods once the linear methods are corrected for the missing persistence bias.

Our second, more substantial, application revisits Boivin, Giannoni and Mihov's (2009) finding that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks (see also Mackowiak, Moench and Wiederholt, 2009). This widely cited finding supports models in which agents choose how much information they acquire because in these models agents respond faster to shocks with a larger variance.[2] While these models may still capture an important aspect of price-setting, we show that Boivin, Giannoni and Mihov's (2009) persistence measure is subject to the missing persistence bias, and that once we correct for it, the responses of sectoral inflation to both types of shocks look very similar. This application illustrates an important point. Results that find persistence measures vary systematically with levels of aggregation should be examined with care, since the differences in estimated speeds of adjustment may be manifestations of the bias.

[2]See standard rational inattention models such as Mackowiak and Wiederholt (2006) and more recent rational inattention/imperfect information hybrids such as Stevens (2016) and Baily and Blanco (2016).

The intuition underlying our main result is best explained in a scenario with three simplifying assumptions: we consider only one agent, shocks faced by this agent are independent, identically distributed (i.i.d.), with zero mean, and the probability that the agent adjusts in any period is constant (and equal to $1-\rho$), as in the discrete time version of the Calvo (1983) model.

Under these assumptions, the agent's adjustments are equal to the accumulated shocks since its previous adjustment. It follows that, if the agent responds in period $t+k$ to a shock that took place in period t, the response must be one-for-one and the agent must not have adjusted in any period between t and $t+k-1$. Therefore, 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 after the shock takes place until the first opportunity to adjust. That is, the average response in $t+k$ is $\rho(1-\rho)^k$, showing that the true, nonlinear, impulse response will be equal to that of a linear AR(1) process with persistence parameter ρ.

Next, consider the impulse response obtained using a linear time-series model. This response depends on the correlations between the agent's actions in different time periods. If the agent did not adjust in one of the periods under consideration, there is no correlation between the amount she adjusted in either period since at least one of the variables entering the correlation is exactly zero. This correlation is also zero if the agent adjusted in both periods 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 positive lags, suggesting immediate adjustment to shocks and therefore no persistence, independent of the true value of ρ. That is, the nonlinear IRF recovers the Rotemberg (1987) result, according to which the aggregate of interest follows an AR(1) process with first-order autocorrelation equal to the fraction of units that remain inactive. However, the linear IRF implies an i.i.d. process that corresponds to the aforementioned AR(1) process when all units adjust in every period and wrongly suggests instantaneous adjustment to shocks, i.e., that $\rho = 0$.

This 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 provide the correlation that allows econometricians 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 by the frequency with which adjustments take place— the faster the estimate converges to the true value of the persistence parameter as the number of agents grows. While idiosyncratic productivity and demand shocks smooth away microeconomic non-convexities and are often cited as a justification for approximating aggregate dynamics with linear models, their presence exacerbates this bias. The fact that, in practice, idiosyncratic uncertainty is many times larger than aggregate uncertainty suggests that the problem of missing aggregate dynamics is prevalent in empirical and quantitative macroeconomic research.

The remainder of the paper is organized as follows. Section 2 presents Rotemberg's (1987) result that justifies using linear time-series methods to estimate dynamics for aggregates with lumpy microeconomic adjustment, as long as there are infinitely many units in the aggregate. Section 3 begins by presenting the missing persistence bias that arises when the number of units considered is finite. Then we describe approaches to correct for this bias. This section concludes by illustrating two implications from the bias for applied researchers. In Section 4 we show that this bias is robust to state-dependent models and strategic complementarities. Section 5 provide a methodology for quantifying the importance of the bias. Section 6 exams two detailed applications and Section 7 concludes. Several appendices follow.

2 LINEAR TIME-SERIES MODELS AND THE CALVO-ROTEMBERG LIMIT

Whether they are trying to identify structural parameters, assess the performance of a calibrated model, or identify a reduced-form characterization of aggregate dynamics, most macroeconomic researchers, at some point, estimate an equation of the form:

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

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

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

We set up 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 she adjusts in period t (the 'reset value' of y). We have:

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

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

From a modeling perspective, lumpy adjustment entails two distinct 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 the first feature of lumpy adjustment that is crucial to generating missing persistence bias.

We therefore start by focusing on a model that has only the first feature of lumpy adjustment, the well-known Calvo model (1983).[3] In this model:

$$\Pr\{\xi_{it} = 0\} = \rho, \qquad \Pr\{\xi_{it} = 1\} = 1 - \rho, \tag{3}$$

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

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

so that ρ represents the degree of *inertia* of Δy_{it}. When ρ is large, the agent adjusts on average by a small fraction of its current imbalance and the expected half-life of shocks is large. Conversely, when ρ is small, the agent reacts promptly to any imbalance. A GE Calvo model with random walk shocks provides a structural interpretation for these reduced form equations (see Appendix E.1.5).

Let us now consider the behavior of aggregates. Given weights w_i, $i = 1, 2, ..., n$, with $w_i > 0$ and $\sum_{i=1}^n w_i = 1$, we define the *effective number of units, N*, as the inverse of the Herfindahl index, $N \equiv 1/\sum_{i=1}^n w_i^2$. We have $N = n$ when all units contribute the same weight to the aggregate ($w_i = 1/n$); otherwise, the effective number of units can be much smaller than the actual number of units.

We write the aggregate at time t, y^N_t, and the corresponding aggregate of reset values, y^{N*}_t as:

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

Technical Assumptions (Shocks)

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

1. The v^A_t's are i.i.d. normal,[4] with zero mean and variance $\sigma^2_A > 0$.
2. The v^I_{it}'s are independent (across units, over time, and with respect to the v^A's), identically distributed normal random variables with zero mean and variance $\sigma^2_I > 0$.
3. The ξ_{it}'s are independent (across units, over time, and with respect to the v^A's and v^I's), identically distributed Bernoulli random variables with probability of success $1-\rho \in (0,1]$. ∎

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

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

[3]In Section 4, we consider state-dependent price models and demonstrate that the mechanisms underling the bias are the same as in the Calvo model.

[4]Normality in Technical Assumptions 1 and 2 is only necessary for the state-dependent results in Section 4.

Taking first differences yields

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

which is analogous to Euler equations derived from a simple quadratic adjustment cost model applied to a representative agent (for the proof, see Appendix D.2).

This is a powerful result that supports the standard practice of treating aggregates with lumpy microeconomic adjustment as if they were generated by a simple linear model. We will show, however, that this approximation can run into problems when motivating the use of VARs for estimating aggregate dynamics. Before doing so, we will close the loop by recovering equation (1) in this setup. Let us momentarily relax Technical Assumptions 1 and 2, allowing for persistence in v_t^A and v_{it}^I, so that the change in the aggregate reset value of y, $\Delta y^{\infty *}$, is generated by:

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

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

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

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

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

Comparing this equation with (1) we conclude that

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

The bias we highlight in this paper comes from a severe downward bias in the (explicit or implicit) estimate of ρ, resulting in an estimate for $a(L)$ that misses significant dynamics. In the next section we simplify the exposition and set $b(L) \equiv 1$, as in the case considered by the Technical Assumptions. We consider the general case in Section 4.

3 THE MISSING PERSISTENCE BIAS

The effective number of units, N, in any real world aggregate is not infinite. In this section, we investigate whether N is large enough that the limit result provides a good approximation. Specifically, we ask whether estimating (6) with an effective number of units equal to N instead of infinity yields a consistent (as T goes to infinity) estimate of ρ, when the true microeconomic model is described by (2) and (3). The following proposition answers this question by providing an explicit expression for this bias as a function of the parameters characterizing adjustment probabilities and shocks (ρ,

σ_A and σ_I) and N.

Proposition 1 (Aggregate Bias)

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

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

Let T denote the length of the time series. Then, under the Technical Assumptions:

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

with

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

Proof See Appendix D. ∎

Letting N tend to infinity in (8) we have that $K/(1+K)$ tends to one and obtain Rotemberg's (1987) result. Yet here we are interested in the value of $\hat{\rho}^N$ before the limit is reached and how structural parameters affect the magnitude of the missing persistence bias. Examining equation (9) reveals the following simple comparative statics: this bias is decreasing in the effective number of units (N), the fraction of agents that adjust each period (1-ρ) and the size of aggregate shocks (σ_A), and increasing in the size of idiosyncratic shocks (σ_I).

3.1 What is Behind this Bias and Slow Convergence?

Next we turn to the intuition behind the proof of the proposition. We do this in two steps. We first describe the genesis of this bias, which can be seen most clearly when $N = 1$. We then show why, for realistic parameter values, the extreme bias identified for $N = 1$ vanishes slowly as N grows.

3.1.1 Genesis of this Bias

To understand where the bias comes from, we first note that Proposition 1 implies that when $N = 1$,

$$\text{plim}_{T\to\infty}\hat{\rho}^1 = 0. \tag{10}$$

That is, a researcher using a linear model to infer the speed of adjustment from the series for one unit will conclude that adjustment is infinitely fast, independent of the true value of ρ.[5] To see why

[5] 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 this bias step-by-step.

this is so, we write

$$\text{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) = \text{E}[\Delta y_t \Delta y_{t-1}] = \sum_{i=0}^{1}\sum_{j=0}^{1} \text{E}[\Delta y_t \Delta y_{t-1} | \xi_t = i, \xi_{t-1} = j]\,\text{Pr}(\xi_t = i, \xi_{t-1} = j), \tag{11}$$

where we used that $\text{E}[\Delta y_t] = 0$ in the first step (see Proposition A.6 in the appendix). We argue that each of the four terms in the sum is equal to zero. The three terms where the unit did not adjust in at least one of the periods will be zero because either Δy_t or Δy_{t-1} (or both) are equal to zero. This leaves only one term where $\Delta y_t \Delta y_{t-1}$ can be different from zero, the case when the unit adjusts both in t and in $t-1$. Yet in this case

$$\Delta y_t \Delta y_{t-1} \;=\; \Delta y_t^* (\Delta y_{t-1}^* + \ldots + \Delta y_{t-s-1}^*),$$

where s denotes the number of periods with inaction prior to adjustment in $t-1$. And since shocks in non-overlapping periods are independent, the expectation of the above product is also zero.

3.1.2 Slow Convergence

To understand what is behind slow convergence, we express $\hat{\rho}$ in terms of four covariance terms:

$$\text{plim}_{T\to\infty} \hat{\rho}^N = \frac{\text{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)}{\text{Var}(\Delta y_t^N)} \;=\; \frac{\sum_i w_i^2 r_{ii}(1) \;+\; \sum_{i\neq j} w_i w_j r_{ij}(1)}{\sum_i w_i^2 r_{ii}(0) \;+\; \sum_{i\neq j} w_i w_j r_{ij}(0)},$$

where $r_{ij}(k) \equiv \text{Cov}(\Delta y_{it}, \Delta y_{j,t-k})$ denotes the covariance of adjustments of units i and j at k lags.

Because units enter symmetrically in the Technical Assumptions, there are only two different covariance functions in the above expression, an auto-covariance function when $i = j$ and a cross-covariance function when $i \neq j$. We denote the former by $r_a(k)$ and the latter by $r_c(k)$. Using that $N = 1/\sum_i w_i^2$ and $\sum_i w_i = 1$, the above expression then simplifies to

$$\text{plim}_{T\to\infty} \hat{\rho}^N \;=\; \frac{\frac{1}{N} r_a(1) + \frac{N-1}{N} r_c(1)}{\frac{1}{N} r_a(0) + \frac{N-1}{N} r_c(0)}. \tag{12}$$

The expressions derived in Appendix D and (11) imply that

$$r_a(0) = \sigma_I^2 + \sigma_A^2, \qquad r_c(0) = \frac{1-\rho}{1+\rho}\sigma_A^2. \qquad r_a(1) = 0, \quad r_c(1) = \rho\frac{1-\rho}{1+\rho}\sigma_A^2. \tag{13}$$

It follows from (12) and (13) that linear time-series models use a combination of auto- and cross-covariance terms to estimate the microeconomic persistence parameter. Inaction biases the auto-covariance terms toward infinitely fast adjustment as $\text{plim}_{T\to\infty} \hat{\rho}^N = 0$ when $N = 1$. The speed with which linear time-series models recover the true value of ρ depends on the extent to which the cross-covariance terms play a dominant role. Since these terms use the common components in the adjustment of different units in consecutive periods to recover ρ, their contribution when estimating ρ will be smaller when adjustment is less frequent (larger ρ). Also, the covariance terms are proportional to σ_A^2 while the denominator includes a variance term, $r_a(0) = \sigma_I^2 + \sigma_A^2$, that is much larger because micro estimates indicate that idiosyncratic uncertainty is much larger than aggregate uncertainty. These factors explain why convergence can be very slow.

3.2 Bias Correction

This section studies an approach to correct for the missing persistence bias, based on using a proxy for the reset value y^*. In Appendix C we discuss two alternative approaches—one based on an ARMA representation of Δy_t^N and the other based on instrumental variables.

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

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

Equation (14) suggests using a proxy of the shock, Δy^*, to correct for the bias. Since the regressors are orthogonal, from Proposition 1 we have that the coefficient on Δy_{t-1} will be biased downward. By contrast, the true speed of adjustment can be estimated directly from the parameter estimate associated with Δy_t^*, as long as we do *not* impose the constraint that the sum of the coefficients on both regressors must add up to one. Of course, the estimate of ρ will be biased if the econometrician imposes the latter constraint. We summarize these results in the following proposition.

Proposition 2 (Bias with Regressors)

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

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

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

(i) without any restrictions on b_0 and b_1:

$$plim_{T\to\infty} \hat{b}_0 = \frac{K}{1+K}\rho, \tag{16}$$

$$plim_{T\to\infty} \hat{b}_1 = 1-\rho; \tag{17}$$

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

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

Proof See Appendix D. ∎

Proposition 2 conveys the general message that constructing a proxy for the reset variable y^* can be very useful when estimating the dynamics of a macroeconomic variable with lumpy microeconomic adjustment. This proposition also suggests not imposing constraints that hold only when $N = \infty$. This proposition is at the center of the Application we consider in Section 5.2.

Proposition 2 is also useful for explaining why the missing persistence bias is not a particular case of an omitted variable bias. The omitted variable bias occurs when a regressor that is correlated with other regressors is not included in the estimation equation. This omission biases the estimates for the regressors that were included. This is not the case in our setting, since Δy_t^* is orthogonal to Δy_{t-1} and the coefficient for Δy_t is still biased after we include Δy_t^*. Nonetheless, the coefficient for Δy_t^* allows us to estimate the speed of adjustment.

3.3 Implications for Empirical Researchers

This section studies the implications of the missing persistence bias for two important tools in the applied macroeconomist's toolkit: the estimation of impulse response functions and simulation based estimation. We give two warnings and provide two remedies.

3.3.1 Estimating Impulse Response Functions

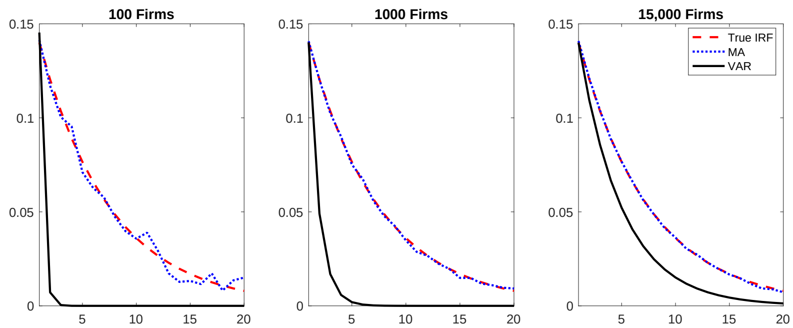
There are two main methods for estimating impulse response functions (IRFs) to an identified structural shock (Ramey 2016). First, the "VAR approach" estimates a vector autoregression and uses the estimated system of equations to compute the IRF. Second, the "MA approach", closely related to Jorda's (2005) local projection method, regresses the series of interest on k lags of the structural shocks. Estimated coefficients from the MA approach then correspond to the elements of the IRF.

These two methods are equivalent for linear models with infinitely long samples (see Christiano, Eichenbaum and Evans (1999) for details), yet the VAR approach is more common in practice. IRF estimates obtain via the MA approach are less popular because they are are less precise and can behave erratically, since this approach imposes no restriction on the shape of the IRF, in contrast to a low order VAR. Despite these limitations, the MA approach has some merits. Ramey (2016) argues that the MA approach is more robust when the estimated VAR is misspecified, which might happen if the true dynamics are non-linear. In this case, the VAR approach will compound these specification errors at each horizon of the IRF. We highlight a second reason to prefer the MA approach: it is robust to the missing persistence bias.

Consider the following simple example. A policymaker wishes to estimate the response of inflation to a monetary policy shock and the adjustment of prices is lumpy. We explore this scenario using a standard version of the Calvo model.[6] The only novelty is that we vary the number of un-

[6]For details on the model see the Calvo random walk calibration reported in Appendix F.2.

Figure 1: RESPONSE OF INFLATION TO A NOMINAL SHOCK (CALVO MODEL)



This figure shows the IRF of inflation to a nominal shock computed in three ways: 1) Using the analytical expression for the IRF: $(1-\rho)\rho^k$ (red-dashed line); 2) Using our MA methodology (light blue dotted-line) 3) Using our VAR methodology (black solid line). All are the average IRFs across 100 simulations.

derlying agents in the economy instead of assuming a continuum.

Each of the panels in Figure 1 considers a different number of agents (100, 1,000 and 15,000) and shows three curves. The first curve (red dashed line) is the theoretical IRF to a nominal shock at k lags, which is equal to $(1-\rho)\rho^k$ for this model, where $(1-\rho)$ is the frequency of price adjustment. The second curve (blue dots) reports the IRF obtained via the MA approach, averaging the estimates from 100 simulations. This curve is almost indistinguishable from the true IRF. In contrast, the third curve (black solid line) reports the results obtained with the VAR approach, also averaging 100 simulations. These estimates are severely downward biased, particularly for small N. The estimated IRF using the VAR approach is always below the true response. Thus researchers using the VAR approach will infer much faster adjustment to nominal shocks than the true value. By contrast, the MA approach is robust to the missing persistence bias.

Proposition 2 is useful for understanding why the MA approach for estimating the IRF is immune to the missing persistence bias while the AR approach is not. Let Δy_t^N denote aggregate inflation when the effective number of units is N and Δy_t^* is the nominal shock, and assume the true model for a continuum of firms is

$$\Delta y_t^\infty = \sum_{k=1}^{p} a_k \Delta y_{t-k}^\infty + c\Delta y_t^*.$$

The VAR approach then estimates:

$$\Delta y_t^N = \sum_{k=1}^{p} a_k \Delta y_{t-k}^N + b\Delta y_t^{*N} + e_t,$$

and the same logic underlying Proposition 2 explains why the estimates of the a_k's are biased. The bias arises because the correlation between Δy_t^N and Δy_{t-j}^N is smaller, in absolute value, than the correlation of Δy_t^∞ and Δy_{t-j}^∞ for all $j \geq 1$.[7]

The MA approach consists in estimating

$$\Delta y_t^N = \sum_{k=0}^{p} I_k \Delta y_{t-k}^* + e_t.$$

As with Proposition 2, coefficients for Δy_{t-k}^* will not be biased.

The general message from this simple exercise is that, when regressing an aggregate with lumpy micro adjustment, including lags of the dependent variable generally leads to biased estimates, while using projection methods that use proxies for the shocks does not.[8] We build on this insight in the application we consider in Section 6.2.

3.3.2 Simulation Based Estimators

Simulation based estimators are often used because they do not require solving the model explicitly and it suffices to be able to generate data from the model. One popular approach is indirect inference, where parameters are chosen to minimize the distance between data moments and moments generated by an auxiliary model. Under mild assumptions, this approach identifies the structural parameters of interest (Smith, 2008).

While indirect inference has many virtues, this methodology must be applied with care if the missing persistence bias is present. Table 1 illustrates this point with a simple Monte Carlo simulation that builds on our previous Calvo model. Consider an applied researcher who wants to estimate the frequency of adjustment (the structural parameter) by SMM using the impulse response function of inflation to a nominal shock as the auxiliary model.[9] This IRF is a sensible choice since the k^{th} element of the IRF is equal to $\rho^k(1-\rho)$.[10] Assume that there are 400 price setting firms in the data who all use Calvo pricing with the same frequency of adjustment, $1-\rho$, equal to 0.25. The data moment is the IRF of inflation to a nominal shock computed in this model.

Table 1 shows what happens if the researcher uses a larger number of agents in simulations than are present in the data. The first and second row report SMM estimates, for different sample sizes when both the data and the auxiliary model IRFs are computed using the standard VAR and MA

[7] In Proposition 1 we showed that this ratio is $K/(1+K)$ for $j = 1$. In Appendix A we show that this is the ratio for values of $j \geq 2$ as well.

[8] An interesting application of this observation is estimating the New Keynesian Phillips Curve (NKPC). In Appendix G.4, we show that estimating the NKPC via GMM leads to biased estimates, particularly if $N < 400$, if lagged inflation is in the instrument set (see Gali and Gertler, 1999).

[9] This example was motivated by the classic Christiano, Eichenbaum and Evans (2005) paper. In the language of indirect inference, their auxiliary model is the IRF of eight macroeconomic variables to a monetary policy shock where these IRFs are computed from an identified VAR (the VAR approach from the previous subsection). They then estimate six parameters of their medium scale DSGE model by minimizing the distance between these eight impulse response functions and their counterparts in the model. Similar estimation procedures can be found in Rotemberg and Woodford (1997), Amato and Laubach (2003), Gilchrist and Williams (2000) and Boivin and Giannoni (2006).

[10] Obviously, this is a highly stylized example – in more complicated frameworks this IRF would depend on more than one structural parameter. The example is deliberately kept simple to illustrate the main point.

Table 1: SMM TABLE

Monte Carlo example: matching IRFs by simulated method of moments (SMM)

	Estimator	**Auxiliary model moments** $(1-\hat{\rho})$ Effective number of agents (N) in simulations 400	1,000	4,000	15,000
Data: $N = 400$, $1-\rho = 0.25$	VAR	0.250	0.710	0.820	0.840
	MA	0.250	0.250	0.250	0.250

All rows show the estimated $1-\hat{\rho}$ from the SMM estimation and all results are averages across 100 simulations.

approaches, respectively. Moment weights are calculated optimally; the results are similar if we use proportional weights or the identity matrix.

The first column shows that SMM provides an unbiased estimator of the frequency of adjustment when the researcher's simulation has the same number of firms in the auxiliary model as are in the data, even when using the VAR approach. This supports the folk wisdom that researchers should treat real and simulated data as similarly as possible. The perils of treating them differently are shown in the other three columns of the first row. Since the researcher is using the VAR approach to estimate the IRF, this bias in the auxiliary moment obtained from the "actual" data is much larger than this bias in the simulated data. To reconcile both sets of moments, SMM infers a much faster speed of adjustment than exists in the actual data. For example, if a researcher tried to match this IRF using a simulation with 15,000 firms, she would infer a speed of adjustment of 0.84, even though the speed of adjustment in the actual data is only 0.25. In contrast, the bottom panel shows that no such issue exists if IRFs are estimated by the MA approach.

We conclude that using the MA approach (projection methods) or ensuring that the auxiliary model reflects all known characteristics of the model generating the actual data, including the effective number of agents, are effective safeguards against the missing persistence bias.

4 STATE-DEPENDENT MODELS AND STRATEGIC COMPLEMENTS

The closed-form expressions and simple intuitions we obtained in Section 3 were possible because of the Technical Assumptions from Section 2. In this section we show that this bias is significant under more general assumptions. We focus on two departures from our baseline that are motivated by empirical reality: allowing state-dependent (menu-cost) models (Section 4.1) and allowing agents' decisions to be strategic complements (Section 4.2). In Appendix B we consider three additional extensions: non-zero mean for the aggregate shock v^A, departures from the i.i.d. assumption for shocks and the presence of time to build. We show that the bias remains significant in all cases.

4.1 State-Dependent Models

The Calvo adjustment assumption we made in Section 3 does not capture the state dependency of the likelihood of unit adjustment: units are more likely to adjust when the imbalance is larger. Next we consider models that incorporate this additional element of reality and argue that the intuitions we gave in Section 3 to explain the missing persistence bias also hold for these models.

We begin by noting that the derivation that led to (12) is valid for models with symmetric heterogeneous agents. It follows that $\text{plim}_{T\to\infty}\hat{\rho}^N$ converges to $\rho_c \equiv r_c(1)/r_c(0)$ as N tends to infinity. In the particular case of Calvo adjustments (see Technical Assumption 3), ρ_c is equal to the fraction of inactive firms, yet this is usually not the case for state-dependent models.

The explanation we gave in Section 3.1 for why $\hat{\rho}^N$ is a downward biased estimate of ρ is reflected in (12) in two ways. First, the numerator is biased downward because $r_a(1) = 0$. Second, the denominator is biased upwards because $r_a(0) \gg r_c(0)$, since the former is of order σ_I^2 while the latter is of order σ_A^2 and empirically $\sigma_I \gg \sigma_A$. Next we argue that both these biases are still present in state-dependent models. We assume Technical Assumptions 1 and 2 continue to hold and generalize Technical Assumption 3 to incorporate state-dependent adjustment as follows:

Technical Assumption 4. There exists a function $\Lambda : \mathbb{R} \to [0,1]$, the *adjustment hazard*, such that the state-variable for unit i, x_{it}, evolves according to (18) and the the relation between the state, x_{it}, and adjustment by the unit, y_{it}, follows (19):

$$x_{i,t+1} = (1-\xi_{it})x_{it} + \Delta y^*_{i,t+1}, \tag{18}$$

$$\Delta y_{it} = \xi_{it}x_{it}, \tag{19}$$

where the ξ_{it}'s are independent (across units and over time) Bernoulli random variables with probability of success $\Lambda(x_{it})$. ∎

Technical Assumption 4 covers many well known state-dependent models. The case of a fixed cost of adjusting prices at the microeconomic level, which yields a two-sided *Ss* policy (see, e.g., Barro, 1972), corresponds to $\Lambda(x) = 1$ if $x \notin [s,S]$ and $\Lambda(x) = 0$ otherwise. The case of i.i.d. idiosyncratic shocks to adjustment costs that are drawn from a non-degenerate distribution leads to a smooth adjustment hazard $\Lambda(x)$ that is decreasing for $x < 0$ and increasing for $x > 0$.[11] Calvo adjustments correspond to the case where $\Lambda(x)$ is equal to $1-\rho$, for all x.

The intuition we provided in Section 3 for why the covariance between consecutive adjustments by the same unit, $r_a(1)$, is zero is based on three assumptions: adjustment is lumpy, there are no strategic complementarities, and innovations (the Δy^*) are independent across periods. This intuition does not depend on whether agents' adjustments are determined by an exogenous process (as

[11] See Caballero and Engel (1999) for a detailed discussion of such a model, Dotsey et al. (1999) for an application to prices in a dynamic general equilibrium context, and Caballero and Engel (1993b) for an estimation of a generalized hazard model for prices.

in the Calvo model considered in Section 3) or state-dependent, since in both cases agents fully adjust to all shocks they have faced since they last adjusted.[12] It follows that the argument we gave to show that the four terms in the sum (11) are equal to zero also holds for state-dependent models and $r_a(1) = 0$ for these models as well. Also, in Appendix A we show that the expression $r_a(0) = \sigma_I^2 + \sigma_A^2$ we derived in the Calvo case also holds for state-dependent models.

To obtain expressions for the remaining two covariances needed to calculate $\text{plim}_{T\to\infty}\hat{\rho}^N$, $r_a(1)$ and $r_c(1)$, we need additional assumptions. We assume that aggregate shocks are small relative to idiosyncratic shocks and interpret this as meaning that agents only consider idiosyncratic shocks when deciding whether to adjust (see Gertler and Leahy (2008) for a similar assumption).[13] Of course, actual adjustments reflect idiosyncratic and aggregate shocks that have accumulated since the unit last adjusted. We refer to this model as the "small σ_A Ss model."

We show in Proposition A.4 in the appendix that for an aggregate with an infinite number of units,

$$\Delta y_t^{\infty} = \sum_{k\geq 0} \gamma_k v_{t-k}^A, \tag{20}$$

where γ_k, $k \geq 0$, denotes the fraction of units that last adjusted k periods ago. This result has two important consequences. First, it implies that the impulse response function of Δy_t^N with respect to the v^A shocks is $(\gamma_k)_{k\geq 0}$ not only for $N = \infty$ but also for any finite integer N (this follows from Property 2 in Caballero and Engel, 2007). Second, noting that the numerator and denominator of (12) converge to $\text{Cov}(\Delta y_t^{\infty}, \Delta y_{t-1}^{\infty})$ and $\text{Var}(\Delta y_t^{\infty})$ when N tends to infinity, we can obtain expressions for $r_c(1)$ and $r_c(0)$ from (20) (see (22) below). These expressions are of order σ_A^2, as was the case for the Calvo model in Section 3, and for the same reasons we gave there. This is the second ingredient we used in Section 3 to explain the missing persistence bias.

Expression (20) implies that, as with the Calvo model, lumpy micro behavior is smoothed by aggregation and the aggregate with an infinite number of units is equal to a linear function of aggregate shocks. Yet, as with the Calvo model, there is a missing persistence bias for aggregates with a finite number of units, as shown in the following proposition.

Proposition 3 (Aggregate Bias for State-Dependent Models)

Consider the small σ_A Ss model described above. Let T denote the length of the time series and let $\hat{\rho}^N$ denote the OLS estimator of ρ in

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + e_t. \tag{21}$$

[12]The assumption of no strategic complementarities matters here, we consider these in Section 4.2.

[13]This assumption can be rationalized adding a small cost of observing the sum of idiosyncratic shocks that occurred since the unit last adjusted and another small cost of observing the sum of aggregate shocks that took place since the last adjustment to the menu cost of changing prices. When σ_I is sufficiently large and σ_A is sufficiently small, the agent will pay the former cost in every period and the latter cost in no period at all.

Then, under Technical Assumptions 1, 2 and 4,

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

with $\rho_c = \sum_{m\geq 0}\gamma_{m+1}\gamma_m / \sum_{m\geq 0}\gamma_m^2$ *and* $K = (N-1)\sum_{m\geq 0}\gamma_m^2/(\sigma_I^2+\sigma_A^2)$. *We also have:*

$$r_a(1) = 0, \qquad r_a(0) = \sigma_I^2+\sigma_A^2, \qquad r_c(1) = \left(\sum_{m\geq 0}\gamma_{m+1}\gamma_m\right)\sigma_A^2, \qquad r_c(0) = \left(\sum_{m\geq 0}\gamma_m^2\right)\sigma_A^2. \tag{22}$$

Proof See Appendix A. ∎

For the model considered in Section 3 we have $\gamma_k = \rho^k(1-\rho)$ and (22) simplifies to (13). It follows that Proposition 1 is a particular case of Proposition 3.

As discussed in Appendix A.2, for general Ss models where units' adjustments are triggered both by idiosyncratic and aggregate shocks, the aggregate with an infinite number of units satisfies

$$\Delta y_t^\infty = \sum_{k\geq 0} I_k v_{t-k}^A + \text{h.o.t.}, \tag{23}$$

where h.o.t. refers to higher order terms involving products and higher moments of the aggregate shocks. This approximation will be good when σ_A is small relative to σ_I, as is the case in practice.

In contrast with the γ_k in (20), I_k is no longer equal to the fraction of agents that last adjusted k periods ago. For example, since the response of Δy_t^N to a positive v^A shock now includes a response at the extensive margin—units that would have remained inactive without the impulse but increase their price because of it and units that were planing to lower their price but remain inactive because of the impulse—we have that $I_0 > \gamma_0$ (see Caballero and Engel, 2007, for a formal proof).

In Appendix A.2 we use (23), to show that, for small σ_A, Proposition 3 continues to hold, approximately, for general Ss models if we replace γ_k with I_k. This suggests that adjustment will be faster for standard Ss models than for their Calvo and "small σ_A Ss model" counterparts, suggesting that the missing persistence bias is larger for the latter than for the former. The quantitative assessment of this bias in Section 5 confirms this.[14] Despite this difference, we find that this bias is quantitatively significant for the applications in Section 6 for all models we calibrate in Appendix F.

Finally, we note that the results regarding IRF estimation discussed in Section 3 extend directly to state-dependent models (see Appendix A.4 and simulation results in Appendix G.1). Using the VAR approach to estimate IRFs yields biased estimates; using the MA approach does not.

4.2 Strategic Complements

Agents' decision variables are neither strategic complements nor strategic substitutes under the Technical Assumptions from Section 2. This may not be a reasonable assumption, as many authors

[14] Note that this implies that, at least for the purposes considered in this paper, the "small σ_A Ss model" differs in an important way form a standard Ss model with small σ_A.

have argued that strategic complementarities are a central to match the persistence implied by VAR evidence (Woodford, 2003; Christiano, Eichebaum and Evans, 1999, 2005; Clarida, Gali and Gertler, 2000; Gopinath and Itskhoki, 2010).

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

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

with $\phi > \rho$ when prices are strategic complements, and $\Delta \log p^*_t$ follows an ARMA(1,1) process with autoregressive coefficient ϕ and moving average coefficient ρ. Based on this insight, we assess the magnitude of this bias via simulations (see Table 13 in Appendix F.5). In our benchmark model with strategic complementarities, we set $\phi = 0.944$ as Woodford recommends. We find that this bias is larger with strategic complements. For example, when $N = 15{,}000$, the relative error for the estimate of ρ increases from 20% to 41%.

The main reason for the larger relative error is that shocks are more persistent with strategic complementarities: $\hat{\rho}^\infty = \phi$ with $\phi > \rho$. Also, when strategic complementarities are present and agents adjust, they no longer fully adjust to the aggregate shocks that accumulated since the last time they adjusted. This decreases the strength of the mechanism that recovers the speed of adjustment, namely the covariance of adjustments across agents (see Section 3).[15]

5 RELEVANCE OF THE MISSING PERSISTENCE BIAS

In this section, we assess the magnitude of the effect of the missing persistence bias on the estimate of the persistence of the US CPI. We focus on inflation because this is the variable of interest in the applications we consider in Section 6. Since the mean, median and maximum number of effective observations in each of the 66 CPI sectors we consider are 187, 142 and 980 respectively,[16] these values for the effective number of units, N, are of particular interest. Also note that $N < 400$ for 60 out of the 66 sectors.

In Appendix F we report estimates for this bias for 14 models. Our calibrations match five standard moments in the pricing literature and the monthly sampling error of aggregate inflation.[17] We consider the Calvo model from Section 3, the small σ_A Ss model of Section 4 and the standard Ss model. We calibrate both general and partial equilibrium models, and analyze models where y^* follows a random walk or an AR(1) processes.

We consider two measures for this bias: the absolute bias, $\hat{\rho}^N - \rho^\infty$, and the relative bias, $(\hat{\rho}^N -$

[15] There's a countervailing effect because the firm's own-price-change correlation is now positive. Yet the impact of this effect on aggregate inflation quickly decreases as the number of firms grows.

[16] Our definition of sectors is close to a two digit level of disaggregation.

[17] We thank one of the referees for suggesting we match this moment.

$\rho^{\infty})/\rho^{\infty}$, where ρ^{∞} denotes the (theoretical) first-order correlation for an aggregate with an infinite number of units (ρ_c in Proposition 3).

Comparisons across models provide some interesting insights. The Calvo Random Walk and the small σ_A Ss model, which also assumes a random walk for y^*, lead to similar bias estimates. By contrast, this bias is much smaller for the standard Ss model with a random walk. As discussed in Section 4.1, this is because standard Ss models respond faster to shocks due to the extensive margin component in their IRFs.

Even though the magnitude of this bias varies across models, it is substantial for all values of the effective number of units, N, relevant for the applications in Section 6. For $N = 100$ and $N = 400$ this bias is above 67% and 33%, respectively, for all models. At the same time, at the aggregate CPI level ($N = 15{,}000$), this bias is negligible for most models and 20% or more for only two models. The latter assessment is conservative, since this bias is generally larger for alternative measures. For example, at the aggregate CPI level, the estimated average half-life is downward biased by 30% or more for 6 of the 14 models considered in Appendix F.

In contrast with the calibration results reported above, the following proposition provides an estimate for the missing persistence bias that does not require taking a stance on the price-setting model. It only requires two readily available moments for aggregate inflation: the standard deviation and the sampling error.[18]

Proposition 4 (Bias Estimation)

Assume Technical Assumptions 1, 2 and 4 hold and the aggregate of interest involves an effective number of agents N^*. For any positive integer N define $a_N \equiv (N^* - N)/(N^* - 1)$. Let $\hat{\sigma}_{\Delta y}$ denote the sample estimates of the standard deviation and let $\hat{\sigma}_{SE}$ denote the sample error of the aggregate under consideration.

Then, for any positive integer N:

$$\frac{\rho^N - \rho^{\infty}}{\rho^{\infty}} = -plim_{T\to\infty} \frac{\hat{\sigma}^2_{\text{SE}}}{a_N \hat{\sigma}^2_{\text{SE}} + (1 - a_N)\hat{\sigma}^2_{\Delta y}}. \tag{24}$$

It follows that, for $N = N^*$:

$$\frac{\rho^{N^*} - \rho^{\infty}}{\rho^{\infty}} = -plim_{T\to\infty} \frac{\hat{\sigma}^2_{\text{SE}}}{\hat{\sigma}^2_{\Delta y}}. \tag{25}$$

Proof See Appendix A.3. ∎

The first row in Table 2 shows the relative bias estimates obtained with CPI data at different levels of aggregation. We use the above proposition and replace the theoretical moments with their observed values: $\hat{\sigma}_{\text{SE}} = 0.00040$ and $\hat{\sigma}_{\Delta y} = 0.0022$. The bias estimates are close to the median estimate from the 14 models calibrated in Appendix F: for $N = 400$, $1{,}000$ and $4{,}000$ there are seven

[18]As shown in Appendix A.3, the bias estimates that follow are also valid for the k^{th} order autocorrelation, $k \geq 1$.

models with a larger bias and seven models with a smaller bias. The second row in Table 2 reports estimates for $\hat{\rho}^N$ obtained from the actual CPI micro database via bootstrap simulations (see Appendix F for details). These values are close to those obtained with the two-moment estimates.[19]

Table 2: ESTIMATING THE MISSING PERSISTENCE BIAS: INFLATION

Measure	Source	Effective number of agents				
		100	400	1,000	4,000	15,000
Relative bias:	Two moment estimate	−0.838	−0.562	−0.339	−0.114	−0.033
	CPI database (bootstrap)	*-0.845*	*-0.615*	*-0.394*	*-0.124*	*-0.042*
Estimate for $\hat{\rho}^N$:	Three-moment estimate	0.053	0.143	0.216	0.290	0.316
	CPI database (bootstrap)	*0.051*	*0.127*	*0.200*	*0.289*	*0.316*

The first row reports the relative bias for the regression coefficient ρ in (7), for aggregates with different numbers of effective agents, N. Estimates were obtained from (24) with the moment values reported in the main text and $N^* = 15,000$. The second row reports bootstrap estimates from the CPI database for the moments in the first row. The third row reports estimates for ρ^N using an extension of (24) that incorporates a third moment (value of $\hat{\rho}$ when estimating (7) with the entire CPI series) and the fourth row reports the corresponding bootstrap estimates.

The third row reports estimates for ρ^N obtained with an extension of Proposition 4 that uses an additional moment, the observed regression coefficient for the CPI series, $\hat{\rho}^{N^*}$ (see Corollary A.1 in the appendix). And the fourth row reports the corresponding bootstrap estimates. The fit is perfect by construction when $N = 15{,}000$, but it is also very good for other values of N.

Proposition 4 is valid under more general conditions than Technical Assumptions 1, 2 and 4 (see Proposition A.8 in the appendix). In particular, it holds with strategic complementarities, time-to-build and when y^* does not follow a random walk.[20]

We end this section by considering two other macroeconomic variables where lumpy microeconomic adjustment has been well established —employment and investment— and use Proposition 4 to estimate the magnitude of the missing persistence bias for both cases. We use estimates for the sampling error published by the BLS (see Appendix F.4 for details).

Table 3 reports how estimates of this bias derived from (25) varies with the level of aggregation for employment and investment data. Each row reports average bias estimates within the corresponding category. The bias is larger than 50% for employment for highly disaggregated series (NAICS 4+). There is no data at this level of aggregation to obtain an estimate of investment. At the NAICS 3-4 level, the relative bias is greater than 30% for both aggregates: 32% for employment and 35% for the investment-to-capital ratio. This bias is also relevant (approximately 17%) for both variables at the super-sector level (e.g., construction). As was the case for prices with Ss models

[19] Appendix G.6 shows that using the entire bi-monthly CPI sample gives very similar results for comparable Ns.

[20] For this statement to be true, we must either have $r_a(1) = 0$ or reinterpret the relative bias as $(\rho^N - \rho^\infty)/(\rho^\infty - \rho_1)$. Using (12) and the bootstrap estimates of ρ^N at different levels of aggregation, we inferred values of the four underlying covariances and obtained $\rho_a \equiv r_a(1)/r_a(0) = 0.010$, suggesting $r_a(1) = 0$ is a good approximation for the CPI.

Table 3: RELATIVE BIAS IN EMPLOYMENT AND INVESTMENT DATA

Estimating the Relative Bias with $-\hat{\sigma}^2_{SE}/\hat{\sigma}^2_{\Delta y}$: Employment and Investment.

Aggregate	Frequency	Level of Aggregation			
		NAICS 4+	NAICS 3-4	NAICS 2	Aggregate
Employment	Quarterly	-0.534	-0.319	-0.173	-0.0146
Investment	Annual	—	-0.346	-0.170	-0.0176

where the extensive margin plays a role, this bias here is minimal at the aggregate level.

Summing up, the above results suggest that researchers should be mindful of the missing persistence bias when using sectoral inflation, employment and investment data. Furthermore, a simple bias estimate based on two easily available moments, that requires minimal assumptions and no model calibrations or simulations provides a good summary of the bias estimates for the 14 inflation models calibrated in Appendix F.

6 APPLICATIONS

The pricing literature is a natural setting in which to study the relevance of the missing persistence bias because numerous studies over the last decade have shown that prices adjust infrequently.[21] We present two applications using CPI micro data that indicate that this bias is of practical relevance and illustrate how to correct for it using the approach outlined in Section 3.2.

Both applications use the CPI research database, which contains individual price observations for the thousands of non-shelter items underlying the CPI over the sample period 1988:03-2007:12. Prices are collected monthly for all items in New York, Los Angeles and Chicago, so we restrict our analysis to these cities to ensure our sample is representative. The database contains thousands of individual "quote-lines" with price observations for many months. In our data set, an average month contains approximately 13,000-18,000 different quote-lines. Quote-lines are the least aggregated observations possible and correspond to an individual item at a particular outlet. For example, one quote-line collected in the research database is a 16 oz bag of frozen corn at a particular Chicago outlet. We exclude sales and product substitutions from our data set.[22]

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

[22] Here we follow the previous literature. For arguments about why we should exclude sales see Eichenbaum, Jaimovich, and Rebelo (2012) and Kehoe and Midrigan (2016); Bils (2009) discusses problems with including product substitutions.

6.1 Application #1: A Simple Test of the Calvo Model

In an influential paper, Bils and Klenow (2004, henceforth BK) conduct a simple test of the Calvo model using sectoral inflation data. Under the assumptions of the Calvo pricing model considered in Section 3 with $N = \infty$, the persistence of sectoral inflation rates, $\hat{\rho}_s$, estimated using an AR(1) model, is approximately equal to one minus the frequency of price adjustment, $1 - \hat{\lambda}_s$. BK implement this test with the CPI micro data and find that, in all sectors, $\hat{\rho}_s$ is substantially smaller than $1 - \hat{\lambda}_s$.[23] They interpret this as strong evidence against the Calvo model. Our paper suggests a more cautious interpretation. Since price adjustment is lumpy and the sectoral inflation series are constructed from relatively small samples, the missing persistence bias could also explain this empirical result.[24]

We can test whether the missing persistence bias is responsible for BK's finding using the bias correction approach outlined in Section 3.2. For this we follow Bils et al. (2012) and proxy sectoral shocks, v_{st}, with the reset value π_{st}, [25] and estimate

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

Proposition 2 implies, that if we estimate β_s and γ_s in the above equation without imposing any constraints across them, then $\widehat{\gamma}_s$ will be an unbiased estimate of $\hat{\lambda}_s$.

Denote the coefficient on our sectoral reset price inflation measure by $\lambda_s^c = \widehat{\gamma}_s$, where the superindex c stands for "corrected" and define $\lambda_s^{\text{VAR}} = 1 - \widehat{\rho}_s$. To gauge the extent to which the λ_s^c's correct the missing persistence bias, we regress the change in estimated adjustment speed in a given sector, $\lambda_s^c - \lambda_s^{\text{VAR}}$, on the magnitude of this bias, $\lambda_s^{\text{micro}} - \lambda_s^{\text{VAR}}$. That is, since we are in a rare situation where we actually know this bias, we are able to estimate the following equation by OLS:

$$(\lambda_s^c - \lambda_s^{\text{VAR}}) = \alpha + \eta \text{bias}_s + \epsilon_s, \tag{26}$$

with $\text{bias}_s \equiv \lambda_s^{\text{micro}} - \lambda_s^{\text{VAR}}$. Here η is the coefficient of interest as it captures the extent to which our bias correction actually decreases this bias. If the bias reduction is large but unrelated to the magnitude of this bias, the estimated value of α will be large while η won't be significantly different from zero. By contrast, if the bias reduction is proportional to the actual bias, we expect an estimate of η that is significantly positive, taking values close to one if this bias completely disappears.

The first column of Table 4 shows the results. Since the estimated value of η is not statistically different from one and the constant term is close to zero, these results suggest that our bias correction strategy comes very close to eliminating this bias entirely. We interpret this as evidence for the

[23] We have also computed this exercise using the entire bimonthly sample and find results a) similar to our baseline monthly results and b) consistent with our theory: $\hat{\rho}$ is higher in the bi-monthly sample with a mean (median) $\hat{\rho}$ equal to 0.148 (0.106) versus 0.084 (0.06).

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

[25] See Appendix G.5 for details on our implementation of the reset price methodology.

Table 4: MISSING PERSISTENCE BIAS: CROSS-SECTIONAL EVIDENCE

	CPI	Ss	Calvo	CPI	Ss	Calvo
	(Bias Correction)			(Bias reduction)		
η	1.004	1.071	1.023			
	(0.028)	(0.028)	(0.005)			
Frequency				-1.176	-0.257	-1.057
				(0.133)	(0.133)	(0.146)
N				-0.350	0.015	-0.110
				(0.123)	(0.106)	(0.132)
Constant	-0.063	0.042	-0.003	1.003	0.550	0.614
	(0.024)	(0.015)	(0.003)	(0.030)	(0.026)	(0.032)
Observations	66	66	66	66	66	66
R-squared	0.951	0.959	0.998	0.632	0.059	0.493

The first three columns estimate equation (26) with the CPI microdata in a calibrated Ss model and in a calibrated Calvo model, respectively. The main coefficient of interest is η, which captures the extent to which our proposed estimator reduces the missing persistence bias. Columns 4-6 document how the magnitude of this bias, measured by the gap between the VAR implied frequency and the true frequency of adjustment, $\lambda_s^{\text{VAR}} - \lambda_s^{\text{micro}}$, varies across sectors with observables (the frequency of adjustment and the number of effective observations), which Proposition 1 suggests should be related to the magnitude of this bias.

empirical relevance of the missing persistence bias in the CPI micro data.

Next, we conduct the same regressions in calibrated multi-sector Ss and Calvo models. These multi-sector models provide a useful laboratory to test, in a controlled setting, whether the missing persistence bias is relevant and whether our bias correction approach works.[26] The results are reported in columns 2 and 3. They show that our bias correction procedure works well in both models. This was expected for the Calvo model, since it satisfies the assumptions in Section 3.1. However, the fact that our approach also works for the Ss case suggests the procedure applies to more general settings.

Columns 4-6 of Table 4 provide further evidence that the missing persistence bias is at work by explicitly examining the comparative statics implied by Proposition 1. In particular, we use cross-sector variation to explore how the magnitude of the bias, $\lambda_s^{\text{VAR}} - \lambda_s^{\text{micro}}$, varies with underlying parameters that we can directly measure using sector level microdata:[27] the adjustment frequency and the effective number of observations, N_s. We find evidence that the adjustment frequency and the number of observations are both significantly negatively related to the magnitude of this bias.

[26] Our calibration is standard; see Appendix E.2 for details. Since a crucial element in these calibration is working with the correct number of price setters in each sector, we set the number of effective price-setters in each sector equal to the number of effective price-setters in the relevant sector. In particular, we use item level expenditure weights w_i, $i = 1, 2, .., n$, with $w_i > 0$ and $\sum_{i=1}^n w_i = 1$ within each sector. Then the effective number of units in each sector, N_s, is definied as the inverse of the Herfindahl index.

[27] We de not consider σ_A and σ_I, since these parameters are model-dependent and cannot be measured directly from the CPI database. As a robustness check, we ran a regression of the bias ($\lambda_s^{\text{VAR}} - \lambda_s^{\text{micro}}$) on the sectoral, Calvo-model dependent, measure of K_s and obtained a strongly negative coefficient (t-stat of -7) with an R^2 of 0.47.

Overall, this example shows that this bias is relevant at the sectoral level and that through the use of micro data one can implement our bias correction procedure in practice.

6.2 Application #2: Does Inflation Respond More Quickly to Sectoral Shocks?

Sticky-information and costly observation models imply that agents may respond differently to different shocks. Boivin, Giannoni and Mihov (2009) (henceforth BGM) use BLS micro data and find that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks and interpret this result as evidence in favor of these models. An alternative explanation is that this empirical result—different adjustment speeds to shocks at different levels of aggregation—is due to the missing persistence bias. We explore this possibility and show that the difference in speed of adjustment disappears once we correct for this bias.

We start by briefly explaining BGM's approach, leaving the full details to Appendix G.8. They estimate a factor-augmented vector autoregression (FAVAR) that relates a large panel of sectoral price series, Π_t, to a relatively small number of estimated common factors, C_t, which summarizes macroeconomic forces. Next, they regress each sectoral inflation series on these common factors,[28] denoting the predicted aggregate component, $\lambda_i' C_t$, by π_{st}^{agg}, and the residual that captures the sector-specific component, e_{st}, by π_{st}^{sect}. This methodology decomposes each sectoral inflation series into orthogonal aggregate and sectoral components:

$$\pi_{st} = \lambda_s' C_t + e_{st} = \pi_{st}^{\text{agg}} + \pi_{st}^{\text{sect}}. \tag{27}$$

We can use these components to analyze the response of sectoral prices to macroeconomic and sector-specific shocks by estimating the persistence of these two series. BGM do so using a VAR approach: they fit separate AR(13) processes to the π_{st}^{agg} and π_{st}^{sect} series and measure the persistence of shocks as the sum of the 13 AR coefficients.[29] Table 19 shows that despite using different underlying data, we find similar results to BGM when we implement their methodology in the CPI micro data.[30]

One interpretation of these results is that sectoral prices respond faster to sectoral shocks. However, since the estimation strategy described above regresses a lumpy variable on lags of itself (the "VAR approach") and there are fewer prices underlying the sectoral component, π_{st}^{sect}, relative to the aggregate component, π_{st}^{agg}, BGM's results could also be driven by the missing persistence bias. To determine whether this is the case, we implement an MA methodology below.[31]

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

[29] This is an often used persistence measure and is motivated by the observation that, if there is a lot of persistence in the data, then the sum of the AR coefficients should be close to one. For example, if the underlying microdata were generated by a Calvo model with $N = \infty$, then this sum is equal to one minus the frequency of adjustment.

[30] We report results that assume there are 5 common factors.

[31] In Appendix G.8, we provide simulation results showing that the MA method accurately recovers the true underlying amount of persistence, whereas the VAR methodology implies that inflation responds faster to sectoral shocks.

We need estimates of both aggregate, m_t, and sectoral shocks, x_{st}, for each sector s. We use our sectoral reset price shock measures, v_{st}'s from Section 6.1. In particular, our proxy for aggregate shocks is the first R principal components of the vector of v_{st}'s, V_t. We compute the pure sectoral shock as a residual, x_{st}.[32]

With these aggregate and sectoral shocks in hand, we can easily implement our MA approach. We do this by regressing each sectoral inflation series on distributed lags of the aggregate and sectoral shocks:

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

where $\eta_s^k(L) = \sum_{j\geq 0} \eta_{sj} L^j$ and $\nu_s(L) = \sum_{j\geq 0} \nu_{sj} L^j$ denote lag polynomials. In order to parsimoniously estimate these lag polynomials, we model each $\eta_s^k(L)$ and $\nu_s(L)$ as quotients of two second degree polynomials.[33] This allows us to flexibly approximate a variety of possible shapes for our IRFs while maintaining parsimony.[34] The results we obtain are robust to reasonable variations in the order of these polynomials.[35] Crucially for our procedure, because we have a direct proxy for both shocks, our measures of persistence to these shocks are not susceptible to the missing persistence bias.[36]

We use the expected response time as our measure of persistence. Appendix D.3 provides a formal definition of this measure and shows that it is equal to $\rho/(1-\rho)$ in the AR(1) case considered in Section 3, so that more persistence implies a higher expected response time. We compute the expected response time for each of the R aggregate shocks and summarize the R response times to aggregate shocks by their median. That is, the sectoral persistence measures are defined as

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

The results are shown in Table 5. We report medians for the τ_s^{agg} and the τ_s^{sec}, for 12 possible combinations of the number of principal components (PC) and number of lags (nlags). The interquartile ranges (divided by the square root of the number of sectors) are shown in parentheses. No matter the specification, the estimated average responses to aggregate and sectoral shocks using the MA bias correction procedure outline above are similar. For example, the average across the 12 specifications for the expected response times of sectoral inflation to aggregate and sectoral shocks is 2.39 and 2.48 months, respectively. We conclude that, after correcting for the missing persistence bias, there is no longer evidence that sectoral inflation responds differently to aggregate

[32] We include lags of the aggregate shocks in order to allow for some delay in these shocks propagating up the supply chain. Our results are robust to ignoring them.

[33] We do not have enough data to estimate an unrestricted version of this equation given that we only have 254 observations for each series and R is the number of lags in each lag polynomial coefficient.

[34] We implemented this estimation using the polyest command in Matlab.

[35] This robustness check is shown in Appendix G.8.

[36] The discussion at the end of Section 3.3.1 provides the underpinning for this approach in the simple Calvo setting, see Appendix A.4 for an extension to Ss models.

and sectoral shocks.

Table 5: THE RESPONSE OF SECTORAL INFLATION RATES TO AGGREGATE AND SECTORAL SHOCKS

Median of estimated expected response times to shocks

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

7 CONCLUSION

While many microeconomic actions are infrequent and lumpy, large idiosyncratic shocks map these lumpy microeconomic series into smooth, aggregated counterparts. The presumption, then, is that standard linear time series analyses can be applied to these smooth aggregated time series to gauge their dynamic behavior. The main result of this paper is to qualify and challenge this presumption. While this approach is valid for an infinite number of agents, convergence can be slow, precisely because idiosyncratic shocks are usually large. Moreover, we show that this bias is systematic, leading to faster estimated responses of aggregate time series to aggregate shocks than is actually the case, especially away from the limit with infinitely many agents. We also find that the magnitude of this bias is relevant for sectoral series and may be present in some aggregate series as well.

We propose various procedures to correct for this bias and illustrate their usefulness with two applications. These procedures both include estimates for the shocks among regressors while being careful about which lags of the response variable they include (or avoiding them altogether). In the first application, we show that this bias provides an alternative explanation for the persistence-gap reported in Bils and Klenow's (2004). In the second one, we show that the difference in the speed with which inflation responds to sectoral and aggregate shocks (Boivin et al 2009; Mackoviak et al 2009) disappears once we correct for the missing persistence bias.

References

[1] Alvarez, Fernando, Le Bihan, Herve, and Francesco Lippi, “The real effects of monetary shocks in sticky price models: a sufficient statistic approach," *American Economic Review*, **106** (10), Oct. 2016, 2817–51.

[2] Amato, Jeffery D. and Laubach, Thomas, “Rule-of-thumb behaviour and monetary policy,” *European Economic Review*, **47**(5), 2003, 791–831.

[3] Baily, Isaac and Julio Blanco. “Firm Uncertainty Cycles and the Propagation of Nominal Shocks ." *American Economic Journal: Macroeconomics, Forthcoming*

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

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

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

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

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

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

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

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

[12] Christiano, Eichenbaum and Evans. “Monetary Policy Shocks: What Have We Learned and to What End?." *Handbook of Macroeconomics*, 1999.

[13] Christiano, Eichenbaum and Evans. “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy" *J. of Political Economy*, **113**, 2005, 1-45.

[14] Clarida, Richard, Jordi Gali, and Mark Gertler, “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,” *Quarterly Journal of Economics 115*, 2000, 147-180.

[15] Coibion, Olivier. "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 2012. Vol 4(2), 1–32.

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

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

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

[19] Golosov, Michael and Robert E. Lucas Jr., "Menu Costs and Phillips Curves," *J. of Political Economy*, **115**, 2007, 171–199.

[20] Gilchrist, Simon and John C. Williams, "Putty?Clay and Investment: A Business Cycle Analysis" *Journal of Political Economy*, **108**, 2000, 928–960.

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

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

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

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

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

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

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

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

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

[30] Midrigan, Virgiliu, "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations," *Econometrica*, Vol. 79(4), 2011, 1139–1180.

[31] Ramey, Valerie. "Macroeconomic Shocks and Their Propagation ." *Handbook of Macroeconomics*, 2016.

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

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

[34] Smith, Anthony *Indirect inference*, *The New Palgrave Dictionary of Economics*, 2008.

[35] Stevens, Luminita, "Coarse Pricing Policies", Working paper, 2016.

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

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

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

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

Appendix

A STATE-DEPENDENT MODELS

This appendix presents the results used in Sections 4 and 5 to show the relevance of the missing persistence bias for state-dependent models. Section A.1 extends Proposition 1 in the main text to state-dependent models where the history of idiosyncratic shocks determines whether units adjust or not. This is the "small σ_A Ss model" discussed in Section 4. Section A.2 considers standard state-dependent models, where both idiosyncratic and aggregate shocks determine whether units adjust, and derives some useful approximations for the missing persistence bias. Section A.3 derives an estimate for the missing persistence bias that is valid even when Technical Assumptions 1 and 2 do not hold. Furthermore, no model calibrations or simulations are needed to calculate these estimates since they follow directly from available moments of inflation. Finally, Section A.4 generalizes Proposition 2 to state-dependent models, showing that MA specifications are generally immune to the missing persistence bias while AR-specifications are not.

A.1 Small σ_A Ss Model

We begin with the simplest state-dependent model, namely a symmetric Ss model where agents adjust when their state-variable, x_{it}, takes values outside the inaction range $[-B, B]$, with $B > 0$ given.[37,38]

We assume that aggregate shocks are small relative to idiosyncratic shocks and interpret this as meaning that agents only consider idiosyncratic shocks when deciding whether to adjust (see Gertler and Leahy (2008) for a similar assumption). Of course, actual adjustments reflect idiosyncratic and aggregate shocks that took place since the unit last adjusted. Since we maintain Technical Assumptions 1 and 2, this implies that the state-variable that determines whether an agent adjusts is the sum of idiosyncratic shocks since the agent last adjusted and x_{it} evolves according to:

$$x_{i,t+1} = x_{it} I(|x_{it}| \leq B) + v^I_{i,t+1}, \tag{28}$$

where $I(A)$ is the indicator function of condition A, that is, it is equal to one when A holds and equal to zero otherwise.

The unit's adjustment in period t, Δy_{it}. then satisfies

$$\Delta y_{it} = (x_{it} + v^A_t + v^A_{t-1} + \cdots + v^A_{t-s+1}) I(|x_{it}| > B), \tag{29}$$

where $t - s$ denotes the last time unit i adjusted prior to t.

Next we state the technical assumptions we use in this subsection of the appendix.[39]

[37] This model can be rationalized by assuming firms face a fixed cost of adjusting their nominal price and by approximating the firm's instantaneous profit function in the neighborhood of its maximum by a quadratic function.

[38] The extension to the case of asymmetric Ss models and generalized Ss models like the one considered in Section 4 is discussed at the end of this section and straightforward.

[39] Assumptions 1 and 2 are the same as in the main text. Assumption 3' can be presented in terms of Bernoulli random variables that describe adjustment probabilities, as we did in Section 4.1, thereby stressing the fact that Calvo adjustment is a particular case of the environment we consider in this appendix and that Proposition 1 in the main text is a particular case of the more general Proposition A.5 we derive below.

Technical Assumptions: General Case

Let v_t^A and v_{it}^I denote aggregate and idiosyncratic shocks and where the absence of a subindex i denotes an element common to all units.

We assume:

1. The v_t^A's are i.i.d. normal, with zero mean 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 normal random variables with zero mean and variance $\sigma_I^2 > 0$.

3'. Agents follow symmetric two-sided Ss rules in the state variable x_{it} characterized by (28), with adjustments described by (29).

Invariant Density

Denote by $f(x,t)$ the probability density function (p.d.f.) of the state variable x defined in (29) at time t, immediately before adjustments take place. Since adjustments are triggered only by idiosyncratic shocks, $f(x,t)$ will not depend on the history of aggregate shocks. It follows that there exists an invariant p.d.f., $f(x)$, that describes the distribution of x immediately before adjustments at any point in time. We characterize this p.d.f. next.

Define

$$f_0(x) = n(x; 0, \sigma_I^2), \tag{30}$$

where $n(x;\mu,\sigma^2)$ denotes the p.d.f. of a normal distribution with mean μ and variance σ^2. The sequence of cross-sections $f_1(x), f_2(x), \ldots$ is then defined recursively via

$$f_{k+1}(x) = \alpha_k^{-1} \int_{-B}^{B} n(x-u; 0, s_I^2) f_k(u)\, \mathrm{d}u, \tag{31}$$

where $k \geq 0$ and

$$\alpha_k = \int_{-B}^{B} f_k(x)\, \mathrm{d}x. \tag{32}$$

The p.d.f. $f_0(x)$ describes the state variable (28) of a unit that last adjusted this period, the p.d.f. $f_1(x)$ the p.d.f. of the state of a unit that last adjusted one period ago, and so on. The $f_k(x)$ are strictly positive not only for values of x in the inaction range $[-B, B]$ but also for values outside this range because they incorporate the latest idiosyncratic shock.

Next we show that $f(x)$ can be expressed as a convex combination of the $f_k(x)$:

$$f(x) = \sum_{k \geq 0} \gamma_k f_k(x), \tag{33}$$

with the $f_k(x)$ defined above and the γ_k defined below in terms of the α_k.

Proposition A1 *The invariant p.d.f. of the unit's state variable, $f(x)$, satisfies* (33) *with the $f_k(x)$ defined via* (30)–(32) *and*

$$\gamma_k = \alpha_{k-1}\gamma_{k-1}, \qquad k \geq 1. \tag{34}$$

It follows that for $k \geq 1$ we have

$$\gamma_k = (\alpha_{k-1} \cdot \ldots \cdot \alpha_0)\gamma_0 \tag{35}$$

and imposing $\sum_{k\geq 0}\gamma_k = 1$ *leads to*

$$\gamma_0 = \left\{1 + \sum_{k\geq 0} \Pi_{j=0}^{k}\alpha_j\right\}^{-1}. \tag{36}$$

Proof Denote by $g_0(x) = \sum_{k\geq 0}\gamma_k f_k(x)$ a convex combination of the $f_k(x)$. That is, $\gamma_k \geq 0$ and $\sum_{k\geq 0}\gamma_k = 1$. To show that $g_0(x)$ is the invariant p.d.f. if we choose the γ_k appropriately, we proceed as follows: We subject $g_0(x)$ to adjustments triggered by the two-sided Ss policy under consideration, followed by idiosyncratic shocks, and find conditions on the γ_k so that the resulting p.d.f. is equal to $g_0(x)$.

Following adjustment and the idiosyncratic shock, $f_k(x)$ becomes a linear combination of two p.d.f.s: a density equal to $f_{k+1}(x)$ describing the state if the unit does not adjust, and a density $n(x;0,s_I^2)$ describing the state if the unit adjusts. The weight of the first density is α_k, the weight of the second density is $1-\alpha_k$. It follows that adjustment and idiosyncratic shocks transform $g_0(x)$ into $g_1(x) = \sum_{k\geq 0}\tilde{\gamma}_k f_k(x)$ with $\tilde{\gamma}_k = \alpha_{k-1}\gamma_{k-1}$ for $k \geq 1$ and $\tilde{\gamma}_0 = \sum_{k\geq 0}(1-\alpha_k)\gamma_k$. We therefore have that $g_1(x)$ and $g_0(x)$ are identical (and equal to the invariant density) if and only if $\tilde{\gamma}_k = \gamma_k$ for all k. This is equivalent to imposing (34) and (36). ∎

As usual, $f(x)$ has two interpretations. The first interpretation is the one we gave above: it describes the unconditional distribution of one unit's state variable. It also represents the cross-section of the state variable of a continuum of units immediately before adjustment, at any point in time. The state of an individual unit changes over time but, given the assumption that adjustments are only triggered by the history of idiosyncratic shocks, the cross-section does not depend on the history of aggregate shocks and therefore does not vary over time.

The Stopping Time Connection

Next we establish the connection between the sequence of $(\gamma_k)_{k\geq 0}$ and the distribution of the number of periods between consecutive adjustments by a given unit.

Consider a sequence $Z_0, Z_1, Z_2, \ldots$ of i.i.d. normal random variables with zero mean and variance σ_I^2. Define the sequence of partial sums by $S_n = Z_0 + Z_1 + \ldots + Z_n$, $n \geq 0$. Given $B > 0$ define the random variable

$$\tau = \min\{n : |S_n| > B\}. \tag{37}$$

The random variable τ describes the number of periods between consecutive adjustments by a given unit. If the unit adjusts again immediately we have $\tau = 0$; if it remains inactive one period and adjusts in the next period we have $\tau = 1$, and so on. That is, $\tau = k$ means that after adjusting (and setting $x = 0$) the unit received k shocks $Z_0, \ldots, Z_{k-1}$ such that the sums $S_0, \ldots, S_{k-1}$ were all within the inaction range $[-B, B]$, followed by a shock Z_k that led to $|S_k| > B$ and triggered adjustment.

The random variable τ is a stopping time w.r.t. the sequence of random variables $Z_0, Z_1, Z_2, \ldots$ That is, the event $(\tau = n)$ is completely determined by the random variables $Z_0, Z_1, \ldots, Z_n$. This will prove useful below.

Next we introduce the random variable $S_\tau = \sum_{i=0}^{\tau} Z_i$. This random variable is equal to the unit's adjustment the next time it adjusts. The subindex τ captures that the number of periods between consecutive adjustments is random (and equal to $1+\tau$). Both $\mathrm{E}(S_\tau)$ and $\mathrm{E}(S_\tau^2)$ are of interest in what follows, since they determine $r_a(0) = \mathrm{Var}(\Delta y_{it})$, one of the four covariances needed to calculate the regression coefficient in (7).

It would seem natural to argue that

$$E(S_\tau) = E(\sum_{i=0}^{\tau} Z_i) = [1 + E(\tau)]E(Z_i), \tag{38}$$

$$E(S_\tau^2) = E[(\sum_{i=1}^{\tau} Z_i)^2] = [1 + E(\tau)]\text{Var}(Z_i). \tag{39}$$

The above identities do not hold for any random variable τ, but they do hold when τ is a stopping time. They are known as Wald's First and Second identities and we will use them below.

Denoting the cumulative distribution function of τ by $F_k = \Pr(\tau \le k)$, we have that the probability that the unit has not adjusted after k periods, conditional on not having adjusted after $k-1$ periods, that is the α_k we defined earlier, can be expressed in terms of the F_k as:

$$\alpha_k = \Pr(\tau \ge k+1 | \tau \ge k) = \frac{\Pr(\tau \ge k+1)}{\Pr(\tau \ge k)} = \frac{1 - \Pr(\tau \le k)}{1 - \Pr(\tau \le k-1)} = \frac{1 - F_k}{1 - F_{k-1}}, \quad k \ge 0. \tag{40}$$

It follows that for $k \ge 0$:

$$\Pi_{j=0}^{k} \alpha_j = 1 - F_k, \tag{41}$$

and substituting this expression in (36) leads to

$$\gamma_0 = 1 + \sum_{k \ge 0} (1 - F_k). \tag{42}$$

Substituting (41) and (42) in (35) yields

$$\gamma_k = \frac{1 - F_{k-1}}{1 + \sum_{j \ge 0}(1 - F_j)} = \frac{1 - F_{k-1}}{1 + E(\tau)}, \tag{43}$$

where we used that τ is a non negative random variable and therefore

$$E(\tau) = \sum_{k \ge 0} \Pr(\tau > k) = \sum_{k \ge 0} (1 - F_k).$$

In particular, setting $k = 0$ yields

$$\gamma_0 = \frac{1}{1 + E(\tau)}. \tag{44}$$

The following lemma provides identities involving the γ_k and α_k that will be useful shortly.

Lemma A1 *With α_k and γ_k defined above:*

$$\sum_{k \ge 1} k(\gamma_{k-1} - \gamma_k) = 1, \tag{45}$$

$$\sum_{k \ge 1} \sum_{l \ge 1} (\gamma_{k-1} - \gamma_k)(\gamma_{l-1} - \gamma_l) \min(k, l) = \sum_{m \ge 0} \gamma_m^2, \tag{46}$$

$$\sum_{k \ge 1} \sum_{l \ge 1} (\gamma_{k-1} - \gamma_k)(\gamma_{l-1} - \gamma_l) \min(k-1, l) = \sum_{m \ge 0} \gamma_{m+1}\gamma_m. \tag{47}$$

Proof The proof of (45) follows from

$$\sum_{k \ge 1} k(\gamma_{k-1} - \gamma_k) = \sum_{k \ge 1} (k-1)\gamma_{k-1} + \sum_{k \ge 1} \gamma_{k-1} - \sum_{k \ge 1} k\gamma_k = \sum_{k \ge 1} \gamma_{k-1} = 1,$$

where we used (34) in the first step, and properties of a telescopic sum and $\sum_{k\geq 0}\gamma_k = 1$ in the last step.

Denote by $\mathscr{S}_0$ the sum on the l.h.s. of (46). The sum of terms with $\min(k,l) = m$ is equal to the sum of terms with $k = m$ and $l \geq m$ and the sum of terms with $l = m$ and $k \geq m+1$. Adding $(\gamma_{k-1}-\gamma_k)(\gamma_{l-1}-\gamma_l)$ over these terms, and using the properties of a telescopic sum, yields $\gamma_{m-1}^2-\gamma_m^2$ and therefore

$$\mathscr{S}_0 = \sum_{m\geq 1}(\gamma_{m-1}^2-\gamma_m^2)m = \sum_{m\geq 1}\gamma_{m-1}^2(m-1) + \sum_{m\geq 1}\gamma_{m-1}^2 - \sum_{m\geq 1}\gamma_m^2 m = \sum_{m\geq 0}\gamma_m^2. \tag{48}$$

Next denote by $\mathscr{S}_1$ the sum on the l.h.s. of (47). The terms with $\min(k-1,l) = m$ add up to $s_{m-1}-s_m$ with $s_m = \gamma_m\gamma_{m+1}$. A calculation analogous to (48), with s_m in the place of γ_m^2, then leads to (47). ∎

Denote by l_{it} the last time unit i adjusted as of period t. That is, $l_{it} = 0$ if it adjusts in t; $l_{it} = 1$ if it adjusted in $t-1$ and did not adjust in t, $l_{it} = 2$ if it adjusted in $t-2$ and did not adjust in $t-1$ or t, and so on. We can write Δy_{it} as the sum of its idiosyncratic and aggregate components

$$\Delta y_{it} = \Delta y_{it}^I + \Delta y_{it}^A$$

with

$$\begin{aligned}(\Delta y_{it}^I | l_{it} = k) &= X_{ik} I(|X_{ik}| > B), &(49)\\ (\Delta y_{it}^A | l_{it} = k) &= V_{k,t} I(|X_{ik}| > B), &(50)\end{aligned}$$

where X_{ik} denotes a random variable with probability density $f_k(x)$ defined above and V_{kt} denotes the sum of aggregate shocks since the unit last adjusted

$$V_{kt} = \sum_{k=0}^{l_{it}} v_{t-k}^A.$$

Proposition A2 *With the assumptions and notation introduced above, for any unit i*

$$\begin{aligned}\mathrm{E}(\Delta y_{i,t}) &= 0, &(51)\\ \mathrm{Var}(\Delta y_{i,t}) &= \sigma_I^2 + \sigma_A^2, &(52)\\ \mathrm{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) &= 0. &(53)\end{aligned}$$

Proof We have

$$\mathrm{E}[\Delta y_{it}^I] = \gamma_0 \mathrm{E}[\Delta y_{it}^I | \text{adjust in } t] = \gamma_0 \mathrm{E}[\sum_{k=0}^{\tau} Z_k] = \gamma_0 \mathrm{E}(Z_i)[1+\mathrm{E}(\tau)] = 0,$$

where we used (38). We also have

$$\mathrm{E}[\Delta y_{it}^A] = \sum_{k\geq 0}\gamma_k \mathrm{E}[\Delta y_{it}^A | l_{it} = k] = \sum_{k\geq 0}\gamma_k \mathrm{E}[V_{k,t} I(|X_{ik}| > B)] = \sum_{k\geq 0}\gamma_k (1-\alpha_k)\mathrm{E}[V_{k,t}] = 0.$$

Adding up both expressions proves (51)

To prove (52) we note that:

$$\mathrm{E}[(\Delta y_{it}^I)^2] = \gamma_0 \mathrm{E}[(\Delta y_{it}^I)^2|\text{adjust in } t] = \gamma_0 \mathrm{E}[(\sum_{k=0}^{\tau} Z_k)^2] = \gamma_0 \mathrm{Var}(Z_i)[1+\mathrm{E}(\tau)] = \sigma_I^2,$$

where we used (38) and (44). We also have

$$\begin{aligned}
\mathrm{E}[(\Delta y_{it}^A)^2] &= \sum_{k\geq 0} \gamma_k \mathrm{E}[(\Delta y_{it}^A)^2|l_{it}=k] = \sum_{k\geq 0} \gamma_k \mathrm{E}[(V_{k,t})^2 I(|X_{ik}|>B)] \\
&= \sum_{k\geq 0} \gamma_k (k+1)(1-\alpha_k)\sigma_A^2 = \sigma_A^2 \sum_{k\geq 0}(k+1)(\gamma_k-\gamma_{k+1}) = \sigma_A^2,
\end{aligned}$$

where we used the independence of X_{ik} and V_{kt}, the definition of α_k and (45). Also,

$$\mathrm{E}[\Delta y_{it}^I \Delta y_{it}^A] = \sum_{k\geq 0} \gamma_k \mathrm{E}[\Delta y_{it}^I \Delta y_{it}^A | l_{it}=k] = \sum_{k\geq 0} \gamma_k \mathrm{E}[\Delta y_{it}^I|l_{it}=k]\mathrm{E}[\Delta y_{it}^A|l_{it}=k] = \sum_{k\geq 0}\gamma_k \mathrm{E}[\Delta y_{it}^I|l_{it}=k](1-\alpha_k)\mathrm{E}V_{kt} = 0,$$

where we used that Δy_{it}^I and Δy_{it}^A are independent conditional on the value of l_{it}, and that $\mathrm{E}V_{kt}=0$. Combining the three preceding identities yields

$$\mathrm{E}[(\Delta y_{it})^2] = \mathrm{E}[(\Delta y_{it}^I + \Delta y_{it}^A)^2] = \mathrm{E}[(\Delta y_{it}^I)^2] + 2\mathrm{E}[\Delta y_{it}^I \Delta y_{it}^A] + \mathrm{E}[(\Delta y_{it}^A)^2] = \sigma_I^2 + \sigma_A^2.$$

Finally, the proof of (53) is the same as in the case of Calvo adjustment: The covariance between Δy_{it} and $\Delta y_{i,t-1}$ is zero either because the agent did not adjust in (at least) one of the periods or because adjustments in both periods are independent. ∎

Proposition A3 *With the assumptions and notation introduced above, for two different agents i and j, we have:*

$$\mathrm{Cov}(\Delta y_{i,t}, \Delta y_{j,t}) = \sigma_A^2 \sum_{m\geq 0} \gamma_m^2, \tag{54}$$

$$\mathrm{Cov}(\Delta y_{i,t}, \Delta y_{j,t-1}) = \sigma_A^2 \sum_{m\geq 0} \gamma_{m+1}\gamma_m. \tag{55}$$

Proof From (49), (50) and (51), we have

$$\begin{aligned}
\mathrm{Cov}(\Delta y_{it}, \Delta y_{jt}) &= \mathrm{E}[\Delta y_{i,t}\Delta y_{j,t}] \\
&= \sum_{k\geq 0}\sum_{l\geq 0} \mathrm{E}[\Delta y_{i,t}\Delta y_{j,t}|l_{it}=k, l_{jt}=l]\,\Pr(l_{it}=k, l_{jt}=l) \\
&= \sum_{k\geq 0}\sum_{l\geq 0} \mathrm{E}[(\Delta y_{it}^I+\Delta y_{it}^A)(\Delta y_{jt}^I+\Delta y_{jt}^A)|l_{it}=k, l_{jt}=l]\,\Pr(l_{it}=k, l_{jt}=l) \\
&= \sum_{k\geq 0}\sum_{l\geq 0} \gamma_k\gamma_l \mathrm{E}[\Delta y_{it}^A \Delta y_{jt}^A|l_{it}=k, l_{jt}=l] \\
&= \sum_{k\geq 0}\sum_{l\geq 0} \gamma_k\gamma_l(1-\alpha_k)(1-\alpha_l)\mathrm{E}[V_{kt}V_{lt}] \\
&= \sigma_A^2 \sum_{k\geq 0}\sum_{l\geq 0}(\gamma_k-\gamma_{k+1})(\gamma_l-\gamma_{l+1})\min(k+1,l+1) \\
&= \sigma_A^2 \sum_{k\geq 1}\sum_{l\geq 1}(\gamma_{k-1}-\gamma_k)(\gamma_{l-1}-\gamma_l)\min(k,l) \\
&= \sigma_A^2 \sum_{m\geq 1}\gamma_m^2,
\end{aligned}$$

where we used that Δy^I_{it} and Δy^A_{jt} are independent conditional on l_{it} and l_{jt} in the third step, that $\mathrm{E}[V_{kt}V_{lt}] = \min(k+1, l+1)\sigma_A^2$ in the fifth step and (46) in the last step.

An analogous derivation, using that $\mathrm{E}[V_{k,t}V_{l,t-1}] = \min(k, l+1)\sigma_A^2$ if $k \geq 1$ and 0 otherwise, leads to

$$\mathrm{Cov}(\Delta y_{i,t}\Delta y_{j,t-1}) = \sigma_A^2 \sum_{k\geq 1}\sum_{l\geq 1}(\gamma_{k-1}-\gamma_k)(\gamma_{l-1}-\gamma_l)\min(k-1, l).$$

The proof then concludes by applying (47) to calculate this sum. ∎

The following proposition extends (104) in Rotemberg's equivalence result to the Ss model considered here.

Proposition A4 (Extension of Rotemberg's Result: Small σ_A Ss Model)

With the notation and assumptions introduced above, the aggregate of an infinite number of units, Δy_t^∞, satisfies:

$$\Delta y_t^\infty = \sum_{j\geq 0}\gamma_j v^A_{t-j}. \tag{56}$$

It follows that the IRF of Δy^∞ w.r.t. the v^A shock at k lags is equal to γ_k. Furthermore, this is also the IRF for the aggregate of a finite number of units.

Proof We condition on the history of aggregate shocks at time t: $v^A_t, v^A_{t-1}, v^A_{t-2}, \ldots$ and denote by $(\Delta y_t^\infty | l_t = k)$ the contribution to the aggregate of units that last adjusted k periods ago. We then have:

$$\begin{aligned}
\Delta y_t^\infty &= \sum_{k\geq 0}\gamma_k(\Delta y_t^\infty | l_t = k) \\
&= \sum_{k\geq 0}\gamma_k(1-\alpha_k)V_{kt} \\
&= \sum_{k\geq 0}(\gamma_k - \gamma_{k+1})\sum_{j=0}^{k} v^A_{t-j} \\
&= \sum_{j\geq 0}\sum_{k=j}^{\infty}(\gamma_k - \gamma_{k+1})v^A_{t-j} \\
&= \sum_{j\geq 0}\gamma_j v^A_{t-j},
\end{aligned}$$

where in the second step we used that idiosyncratic shocks reflected in adjustments average to zero because we have an infinite number of units, and in the last step we used the properties of the telescopic sum.

It follows from (56) that the IRF of Δy^∞ w.r.t. the v^A shock at k lags is equal to γ_k. We then have, from Property 2 in Caballero and Engel (2007) that this will also be the IRF for any aggregate consisting of a finite number of units. ∎

Similar to what happens for the Calvo model in Rotemberg Equivalence Result, nonlinearities associated with lumpy adjustment vanish as the number of units tends to infinity for the small σ_A Ss model and the aggregate converges to a distributed lag of aggregate shocks. Also, the impulse response at lag k is equal to the fraction of agents that last adjusted k periods ago. Yet the shape of γ_k admits more general shapes than the geometric decay of the Calvo model.

The following proposition extends Proposition 1 to the state-dependent model studied here:

Proposition A5 (Aggregate Bias for Small σ_A State-Dependent Model)

With the notation and assumptions made above, let $\hat{\rho}^N$ denote the OLS estimator of ρ in

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

Let T denote the time series length. Then, under Technical Assumptions 1, 2 and 3', $plim_{T\to\infty}\hat{\rho}^N$ depends on the weights w_i only through N and

$$plim_{T\to\infty}\hat{\rho}^N = \frac{K}{1+K}\rho_c, \tag{57}$$

with

$$\rho_c \equiv \frac{r_c(1)}{r_c(0)} = \frac{\sum_{m\geq 0}\gamma_{m+1}\gamma_m}{\sum_{m\geq 0}\gamma_m^2}$$

and

$$K \equiv \frac{\sigma_A^2(N-1)\sum_{m\geq 0}\gamma_m^2}{\sigma_I^2+\sigma_A^2}.$$

Proof The proof follows from substituting the expressions obtained in Propositions A2 and A3 for the four covariances in the expression for $\text{plim}_{T\to\infty}\hat{\rho}^N$ derived in (12). ∎

As mentioned in the main text, for the Calvo model considered in Section 3 we have $\gamma_k = \rho^{k-1}(1-\rho)$. Substituting this expression in Propositions A4 and A5 yields the original Rotemberg result (see Appendix D.2) and Proposition 1.

A.2 Standard Ss Models

In Section A.1 we assumed that aggregate shocks play no role in determining when units adjust. We relax this assumption in this section. We also relax the assumption that adjustments follow symmetric Ss policies and consider the generalized Ss models from Section 4. We show that in this more general setting the expressions we derived above for $r_a(0)$ and $r_a(1)$ continue holding. We also find approximate expressions for $r_c(0)$ and $r_c(1)$ that generalize the ones we obtained above.

Assume that Technical Assumptions 1, 2 and 4 hold. We show next that there exists a natural generalization of the stopping time argument that led to Proposition 2.

Consider a sequence $Z_0, Z_1, Z_2, \ldots$ of i.i.d. normal random variables with zero mean and variance $\sigma^2 \equiv \sigma_A^2 + \sigma_I^2$. Define the sequence of partial sums by $S_n = Z_0 + Z_1 + \ldots + Z_n$, $n \geq 0$ and define a sequence of independent Bernoulli random variables, $\xi_1, \xi_2, \xi_3, \ldots$ where the success probability of ξ_n is $\Lambda(S_n)$.

Then the random variable

$$\tau = \min\{n : \xi_n = 1\} \tag{58}$$

describes the number of periods between consecutive adjustments. We note that this variable is a stopping time w.r.t. the sequence of random variables (Z_k, ξ_k), $k \geq 0$. That is, the event $(\tau = n)$ is completely determined by the realizations of the random variables $(Z_0, \xi_0), (Z_1, \xi_1), \ldots, (Z_n, \xi_n)$.

We define the γ_k in terms of the distribution of τ as we did in Section A.1. The following proposition extends Proposition A3 to the more general family of state-dependent models considered here.

Proposition A6 *Assume Technical Assumptions 1, 2 and 4 hold. Then, for any unit i*

$$\begin{aligned} \mathrm{E}(\Delta y_{i,t}) &= 0, & (59) \\ \mathrm{Var}(\Delta y_{i,t}) &= \sigma_I^2 + \sigma_A^2, & (60) \\ \mathrm{Cov}(\Delta y_{it}, \Delta y_{i,t-1}) &= 0. & (61) \end{aligned}$$

Proof To prove (59) and (60) we use Wald identities. Specifically, (59) follows from

$$\mathrm{E}[\Delta y_{it}] = \gamma_0 \mathrm{E}[\Delta y_{it}|\text{adjust in } t] = \gamma_0 \mathrm{E}[\sum_{k=0}^{\tau} Z_k] = \gamma_0 \mathrm{E}(Z_i)[1+\mathrm{E}(\tau)] = 0,$$

with τ defined in (58) and where we used (38).

And (60) follows from:

$$\mathrm{E}[(\Delta y_{it})^2] = \gamma_0 \mathrm{E}[(\Delta y_{it})^2|\text{adjust in } t] = \gamma_0 \mathrm{E}[(\sum_{k=0}^{\tau} Z_k)^2] = \gamma_0 \mathrm{Var}(Z_i)[1+\mathrm{E}(\tau)] = \sigma^2,$$

where we used (39) and (44). Finally, the proof of (61) is the same as the one we provided in Section 3.1.1, see (11). ∎

Having obtained exact expressions for $r_a(1)$ and $r_a(0)$, next we derive approximate expressions for $r_c(1)$ and $r_c(0)$. These approximations assume σ_A is small. Yet, by contrast with the "small σ_A model" studied in Section A.1, in what follows aggregate shocks play a role determining when units adjust.

We begin by noting that using a Volterra series expansion we may write

$$\Delta y_t^{\infty} = \sum_{k\geq 0} I_k v_{t-k}^A + O(\sigma_A^2), \tag{62}$$

where the error term, $O(\sigma_A^2)$, involves higher moments and products of the v_{t-k}^A and therefore has a mean of order σ_A^2. It follows that the aggregate of an infinite number of units can be approximated by a distributed lag of the history of aggregate shocks.

Combining (12) with (62) leads to

$$\begin{aligned} r_c(1) &= \lim_{N\to\infty} \mathrm{Cov}(\Delta y_t^N, \Delta y_{t-1}^N) = \left(\sum_{k\geq 0} I_{k+1} I_k\right) \sigma_A^2 + +O(\sigma_A^4), \\ r_c(0) &= \lim_{N\to\infty} \mathrm{Var}(\Delta y_t^N) = \left(\sum_{k\geq 0} I_k^2\right) \sigma_A^2 + +O(\sigma_A^4), \end{aligned}$$

where both error terms are of order σ_A^4. Combining these expressions and Proposition A.6 proves the following extension of Proposition 5.

Proposition A7 (Aggregate Bias for *Ss* Model)

With the notation and assumptions made above, let $\hat{\rho}^N$ denote the OLS estimator of ρ in

$$\Delta y_t^N = \text{const.} + \rho \Delta y_{t-1}^N + e_t. \tag{63}$$

Let T denote the time series length. Then, under Technical Assumptions 1, 2 and 4, $\text{plim}_{T\to\infty}\hat{\rho}^N$ *depends on the weights w_i only through N and*

$$\text{plim}_{T\to\infty}\hat{\rho}^N \;=\; \frac{K}{1+K}\rho_c + O(\sigma_A^2), \tag{64}$$

with

$$\rho_c \equiv \frac{r_c(1)}{r_c(0)} = \frac{\sum_{m\geq 0} I_{m+1} I_m}{\sum_{m\geq 0} I_m^2}$$

and

$$K \;\equiv\; \frac{\sigma_A^2 (N-1) \sum_{m\geq 0} I_m^2}{\sigma_I^2 + \sigma_A^2}. \quad \blacksquare$$

It is straightforward to see that Propositions 1 and 3 are particular cases of the above result. Yet there are two differences worth noting between Proposition A.7 and the particular cases considered in the main text. First, it provides an approximation that will be good only if σ_A is small. Second, while the cross-covariances in Propositions 1 and 3 can be expressed in terms of the distribution of times between adjustments (the γ_k), this is not the case for the cross-covariances in Proposition A.7 (the I_k). In fact, it follows from Property 5 in Caballero and Engel (2007) that $I_0 > \gamma_0$ for both standard and generalized *Ss* models. Furthermore, the difference between I_0 and γ_0 typically is large, as confirmed by the model calibration results reported in Appendix F, with $I_0 \simeq 3\gamma_0$ being a useful benchmark.

A.3 A General Bias Estimate

Next we prove Proposition 4 in the main text. We derive a more general result that includes this proposition as a particular case.

Our starting point is, once again, equation (12). Defining $\rho_a \equiv r_a(1)/r_a(0)$ and denoting $\rho^N \equiv$ $\text{plim}_{T\to\infty}$ this expression implies

$$\rho^N \;=\; b_N \rho_a + (1-b_N)\rho_c \tag{65}$$

with

$$b_N = \frac{r_a(0)}{r_a(0) + (N-1) r_c(0)}. \tag{66}$$

Since $b_1 = 1$ and $b_\infty = 0$, it follows that $\rho^1 = \rho_a$ and $\rho^\infty = \rho_c$. The intuition is the following one: When the aggregate consists of a single unit, the first-order correlation of the "aggregate" is equal to the first-order autocorrelation of an individual unit. And when the aggregate consists of an infinite number of units, the influence of auto-covariance terms disappears and the first-order autocorrelation of the aggregate equals ρ_c. For values of N in between, ρ^N is a weighted average of ρ^1 and ρ^∞, with weights that decrease with N and only depend on two moments: $r_a(0)$ and $r_c(0)$.

We did not use any of the Technical Assumptions to derive (65). For example, ρ_a could be different from zero (it is negative when aggregate shocks have non-zero mean, see Appendix B), And aggregate and idiosyncratic shocks do not need to follow a random walk. Furthermore, the source of frictions could be adjustment costs, as we consider in this paper, or informational, or a combination of both. The only requirement for (65) to hold is that the Δy_{it} be stationary, that the aggregate of interest be (well approximated by) a (weighted) sum of the corresponding micro variables, and that units enter symmetrically. The latter is needed because we assume that the autocovariance function is the same for all units and the cross-covariance function is the same for any pair of different units.

Rearranging terms in (65) yields:

$$\frac{\rho^{\infty}-\rho^{N}}{\rho^{\infty}-\rho^{1}} = b_N. \tag{67}$$

The l.h.s. of (67) is the relative bias, that is, the quotient of this bias from estimating ρ^{∞} with ρ^N and the bias from estimating ρ^{∞} with ρ^1. The relative bias only depends on two of the four covariances involved: $r_a(0)$ and $r_c(0)$. It follows that using any two moments that are determined by these covariances is enough to obtain an estimate of the relative bias. One possible implementation of this insight is presented next.

Proposition A8 (Relative Bias Estimation: A General Result)

Consider N^* units, assume the Δy_{it} are stationary for $i = 1,2,...,N^*$ and define the aggregate $\Delta y_t \equiv \frac{1}{N^*}\sum_{i=1}^{N^*}\Delta y_{it}$.[40] Assume the auto-covariance function is the same for all units and the cross-covariance function is the same for any pair of units and denote these functions by $r_a(k)$ and $r_c(k)$, respectively. Denote by $\hat{\sigma}_{\Delta y}$ and $\hat{\sigma}_{\mathrm{SE}}$ consistent estimates for the standard deviation and the sampling error of Δy_t, respectively. Let N be any positive integer and define $a_N \equiv (N^*-N)/(N^*-1)$.

Then:

$$plim_{T\to\infty}\hat{\sigma}^2_{\mathrm{SE}} = \frac{r_a(0)}{N^*}, \qquad plim_{T\to\infty}\hat{\sigma}^2_{\Delta y} = \frac{1}{N^*}r_a(0)+\frac{N^*-1}{N^*}r_c(0). \tag{68}$$

It follows that, for any positive integer N:

$$\frac{\rho^{\infty}-\rho^{N}}{\rho^{\infty}-\rho^{1}} = plim_{T\to\infty}\frac{\hat{\sigma}^2_{\mathrm{SE}}}{a_N\hat{\sigma}^2_{\mathrm{SE}}+(1-a_N)\hat{\sigma}^2_{\Delta y}}. \tag{69}$$

In particular, if $N = N^*$:

$$\frac{\rho^{\infty}-\rho^{N^*}}{\rho^{\infty}-\rho^{1}} = plim_{T\to\infty}\frac{\hat{\sigma}^2_{\mathrm{SE}}}{\hat{\sigma}^2_{\Delta y}}. \tag{70}$$

Proof The proof boils down to deriving the two expression in (68).

Sampling error estimates for Δy_t are obtained by bootstrapping estimates of the aggregate of interest at time t, that is, by considering random samples of size N^*, calculating aggregate inflation for each sample and then setting the sampling error equal to the standard deviation of the bootstrap estimates. Since samples are random, the variance of each Δy_{it} that is sampled will be equal to $r_a(0)$ and it follows that

$$\mathrm{plim}_{T\to\infty}\hat{\sigma}^2_{\mathrm{SE}} = \frac{r_a(0)}{N^*}. \tag{71}$$

The Ergodic Theorem implies that the time-series and cross-section variances of Δy_t will be the same. The data moment we observe is the former, the moment that is easy to express in terms of

[40] For ease of exposition, we assume equal weights across units.

the covariances is the latter. Indeed, the numerator in (12) is equal to this variance and therefore

$$\text{plim}_{T\to\infty}\hat{\sigma}^2_{\Delta y} = \frac{1}{N^*} r_a(0) + \frac{N^*-1}{N^*} r_c(0). \tag{72}$$

From (71) and (72) we have:

$$r_a(0) = N^* \text{plim}_{T\to\infty}\hat{\sigma}^2_{\Delta y}, \tag{73}$$

$$r_c(0) = \frac{N^*}{N^*-1}\text{plim}_{T\to\infty}(\hat{\sigma}^2_{\Delta y} - \hat{\sigma}^2_{\text{SE}}). \tag{74}$$

Also, given any positive integer N we have that (66) implies

$$b_N = \frac{r_a(0)}{r_a(0) + (N-1) r_c(0)}.$$

Substituting (73) and (74) in the above expression and using (67) then leads to (68) and (69) and completes the proof. ∎

Corollary A1 *Under the assumptions of Proposition A 8, suppose that the researcher has access to a third moment, namely a consistent estimate $\hat{\rho}^{N^*}$ for ρ^{N^*}. Then, for any positive integer N:*

$$plim_{T\to\infty}\frac{(1-a_N)\hat{\sigma}^2_{\Delta \text{y}}}{a_N\hat{\sigma}^2_{\text{SE}} + (1-a_N)\hat{\sigma}^2_{\Delta y}}(\hat{\rho}^{N^*} - \rho^1) = \rho^N - \rho^1. \tag{75}$$

In particular, if $\rho^1 = 0$,

$$plim_{T\to\infty}\frac{(1-a_N)\hat{\sigma}^2_{\Delta \text{y}}}{a_N\hat{\sigma}^2_{\text{SE}} + (1-a_N)\hat{\sigma}^2_{\Delta y}}\hat{\rho}^{N^*} = \rho^N. \tag{76}$$

Proof Since

$$\frac{\rho^1 - \rho^N}{\rho^\infty - \rho^1} = \frac{\rho^\infty - \rho^N}{\rho^\infty - \rho^1} - 1,$$

it follows from (69) that

$$\frac{\rho^1 - \rho^N}{\rho^\infty - \rho^1} = \text{plim}_{T\to\infty}\frac{(1-a_N)(\hat{\sigma}^2_{\text{SE}} - \hat{\sigma}^2_{\Delta y})}{a_N\hat{\sigma}^2_{\text{SE}} + (1-a_N)\hat{\sigma}^2_{\Delta y}}.$$

In particular, since $a_N = 0$, when $N = N^*$ the above expression simplifies to

$$\text{plim}_{T\to\infty}\frac{\rho^1 - \hat{\rho}^{N^*}}{\rho^\infty - \rho^1} = \text{plim}_{T\to\infty}\frac{\hat{\sigma}^2_{\text{SE}} - \hat{\sigma}^2_{\Delta \text{y}}}{\hat{\sigma}^2_{\Delta y}}.$$

Taking the quotient of the two preceding expressions yields (75). ∎

A.4 Higher order correlations and regressors

In this section we extend Proposition 2 in the main text, that provides a strategy to obtain estimates that are not affected by the missing persistence bias, to state-dependent models. This proposition is at the heart of the empirical strategy we use in the applications in Section 6 to estimate the true speed of adjustment. The derivation that follows is exact for the small σ_A Ss model discussed in Section A1 and a good approximation when σ_A is small for the general Ss models considered in Section A.2.

For $k \geq 1$ and $m \geq 1$ denote by ρ_k^m the theoretical k-th order correlation for an aggregate with m effective units. For simplicity, we consider the case where $\rho_k^1 = 0$ for all k.[41] A derivation analogous to the one that led to (67) can be used to show that, for $k \geq 2$:

$$\frac{\rho_k^\infty - \rho_k^N}{\rho_k^\infty} = b_N, \tag{77}$$

That is, on average, the relative bias for the k-th order correlation of an aggregate with N units is the same as for the corresponding first-order autocorrelation. The missing persistence bias shrinks the estimates for all correlations toward zero in the same proportion.

Next we use this result to show how to use a proxy for the aggregate shock to obtain estimates for the speed of adjustment that are immune to the missing persistence bias, thereby extending Proposition 2 to state-dependent models.

Assume, the researcher has observations of the aggregate shock, v_t^A, and wants to decide between estimating an autoregressive process:

$$\Delta y_t^N = \text{const.} + \sum_{k=1}^{L} b_k \Delta y_{t-k}^N + c_0 v_t^A + e_t \tag{78}$$

and a moving average process

$$\Delta y_t^N = \text{const.} + \sum_{j=0}^{M} c_j v_{t-j}^A + e_t. \tag{79}$$

If $N = \infty$ and the number of lags in both regressions are large enough, both approaches are equivalent in theory and estimating (78) is often more efficient, since fewer parameters are needed to obtain a good fit.

Yet when N is finite and the missing persistence bias is significant, (77) implies that (78) will lead to biased estimates. Furthermore, since all correlations are biased toward zero, the estimated process will be closer to an i.i.d. process than the process with an infinite number of units, implying that the implied speed of adjustment will be faster than the true speed.[42]

By contrast, since the v_t^A are i.i.d., estimating (79) will lead to estimates for the c_j that are proportional to the covariance of Δy_t^N and v_{t-j}^A. And since $\text{Cov}(\Delta y_{it}, v_{t-j}^A)$ is the same for all units, we have that $\text{Cov}(\Delta y_t^N, v_{t-j}^A) = \text{Cov}(\Delta y_{it}, v_{t-j}^A)$ for all i and N and therefore also for $N = \infty$. It follows that estimating (79) and then setting $\widehat{\text{IRF}}_j = \hat{c}_j$ will lead to unbiased estimates for the true impulse response function.

[41] This holds for the model considered in Section 3 and the models considered in this appendix. It does not hold when aggregate shocks have a non-zero mean.

[42] Recall that an i.i.d. process implies infinitely fast responses to shocks.

B EXTENSIONS

B.1 Non-Zero Drift

In the main text we assumed that aggregate shocks v^A have zero mean. Here we relax this assumption and show that this bias is larger when we allow for a non-zero mean.

We consider the Calvo model from Section 3 but allow for a non-zero mean for aggregate shocks, μ_A. In Appendix D we derive explicit expressions for the four covariances involved in the calculation of the regression coefficient $\hat{\rho}^N$ (see in (7) and (12)). Two coefficients remain unchanged: $r_c(1)$ and $r_c(0)$. The other two coefficients become:

$$r_a(0) = \sigma_A^2 + \sigma_I^2 + \frac{2\rho}{1-\rho}\mu_A^2, \qquad r_a(1) = -\rho\mu_A^2.$$

It follows from the above expressions and (12) that for given values of σ_A, σ_I, ρ and N, this bias will be larger if $\mu_A \neq 0$, for two reasons. First, $r_a(1)$ now is negative instead of zero, which leads to a smaller numerator in (12). Second, $r_a(0)$ now is larger, which leads to a larger denominator in (12).

To understand the impact of the drift on convergence, we must explain why the covariance between Δy_t and Δy_{t-1} for a given unit is negative when $\mu_A \neq 0$ and why the variance term increases with $|\mu_A|$. To provide the intuition for the negative covariance, assume $\mu_A > 0$ (the argument is analogous when $\mu_A < 0$) and note that the unconditional expectation of Δy_t is equal to μ_A, which corresponds to expected adjustment when the unit adjusts in consecutive periods (the proof follows directly from Wald's First Identity, see (38)). The expected adjustment when adjusting after more than one period is larger than μ_A. It follows that a value of Δy_t above average indicates that it is likely that the agent did not adjust in $t-1$, implying that Δy_{t-1} probably is smaller than average. Similarly, a value of Δy_t below average indicates that probably the agent adjusted in period $t-1$, and Δy_{t-1} is likely to be larger than average in this case.

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

B.2 Adding Smooth Adjustment (Time-to-Build)

The setting is that of Section 3, yet we assume that, in addition to the infrequent adjustment pattern described throughout the paper, 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., Woodford (1999)). Our main result here is that the econometrician estimating a linear ARMA process —a Calvo model with additional serial correlation— will only be able to extract the gradual adjustment component but not the source of sluggishness from the infrequent adjustment component. That is, again, the estimated speed of adjustment will be too fast, for exactly the same reason as in the simpler model.

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

$$\Delta y_t = \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + (1 - \sum_{k=1}^{K} \phi_k) \Delta \tilde{y}_t. \tag{80}$$

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

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

$$\Delta \tilde{y}_t = \rho \Delta \tilde{y}_{t-1} + v_t. \tag{81}$$

Replacing (81) into (80), yields the following linear equation in terms of the observable, Δy_t:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t, \tag{82}$$

with

$$\begin{aligned} a_1 &= \phi_1 + \rho, \\ a_k &= \phi_k - \rho\phi_{k-1}, \qquad k = 2, ..., K, \\ a_{K+1} &= -\rho\phi_K, \end{aligned} \tag{83}$$

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

We show next that the econometrician will miss the source of persistence stemming from ρ.

Proposition A9 (Omitted Source of Sluggishness)

Let all the assumptions in Proposition 1 hold, with $\tilde{y}$ in the role of y. Also assume that (80) 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 (82).

Then:

$$\begin{aligned} \text{plim}_{T\to\infty} \hat{a}_k &= \phi_k, \qquad k = 1, ..., K, \\ \text{plim}_{T\to\infty} \hat{a}_{K+1} &= 0. \end{aligned} \tag{84}$$

Proof See Appendix D. ∎

Comparing (83) and (84) we see that the proposition simply reflects the fact that the (implicit) estimate of ρ is zero.

B.3 Relaxing the Assumption that Shocks are I.I.D.

In Section 3 we assumed that shocks are i.i.d. This is the assumption made by Woodford (2003, sect. 3.2) for nominal output and by Bils and Klenow (2004) for marginal costs. Other authors allow for persistence in shocks. For example Midrigan (2011) considers an autoregressive process for money growth with a persistence parameter of 0.61.

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

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

with ε_t proportional to v_t^A and β denoting the agent's discount factor.[43]

We assume the researcher knows ϕ and β and therefore estimates the fraction of inactive firms, ρ, from:

$$(\Delta y_t^N - \phi\Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi\Delta y_{t-2}^N) + e_t - \beta\phi\rho e_{t-1}. \tag{85}$$

Table 6: SLOW CONVERGENCE WHEN Δy^* FOLLOWS AN AR(1)

Estimated $\hat{\rho}^N$: Δy^* follows an AR(1)

	Effective number of agents (N)					
ϕ	100	400	1,000	4,000	15,000	True
0.00	0.022	0.083	0.182	0.447	0.690	0.860
0.10	0.000	0.001	0.062	0.380	0.680	0.860
0.20	0.000	0.000	0.001	0.283	0.662	0.860
0.30	0.000	0.000	0.000	0.112	0.625	0.860
0.40	0.000	0.000	0.000	0.000	0.566	0.860
0.50	0.000	0.000	0.000	0.000	0.419	0.860

This table reports estimates for ρ in (85), obtained via maximum likelihood, with β and ϕ known and imposing $\rho \geq 0$. Estimates based on 100 simulations with a series of length $T = 238$ each, as in Table 9. Parameters (monthly pricing data): $\rho = 0.86$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\sigma_I = 0.0616$, $\beta = 0.96^{1/12}$.

Table 6 shows the average estimate of ρ in (85) obtained via 100 simulations. Since the researcher knows ϕ and β, the only source of bias is that the researcher ignores the fact that because the actual aggregate considers a finite number of agents, using the linear specification valid for an infinite number of agents will bias the estimated speed of adjustment upwards.[44]

It follows from Table 6 that this bias is generally larger when the innovations of Δy^* are positively correlated than in the i.i.d. case, the increase can be large when N is small. For example, for $\phi = 0.10$ and $N = 1,000$, the estimated value of the inaction parameter ρ is only one-third of the value estimated when shocks follow a random walk: 0.062 vs. 0.182. For the same value of N and larger values of ϕ estimates of ρ are equal to zero and the researcher infers an infinite speed of adjustment if she ignores the missing persistence bias.

C ADDITIONAL BIAS CORRECTION METHODS

In the main text we studied an approach to correct for missing persistence bias using a proxy for y^*, which is the approach we used in Section 6. Here we provide two additional approaches.

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

[44] Simulations show that this bias disappears if we estimate $(\Delta y_t^N - \phi\Delta y_{t-1}^N) = \text{const.} + \rho(\Delta y_{t-1}^N - \phi\Delta y_{t-2}^N) + e_t - \gamma_1 e_{t-1} - \gamma_2 e_{t-2}$ with no constraints on γ_1 and γ_2. This suggests that the random walk assumption can be relaxed in Proposition A.10. We thank Juan Daniel Díaz for this insight.

C.1 ARMA Correction

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

Proposition A10 (ARMA Representation)

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

$$\Delta y_t^N = \rho \Delta y_{t-1}^N + (1-\rho)[\varepsilon_t - \theta \varepsilon_{t-1}], \tag{86}$$

where ε_t is an i.i.d. innovation process and $\theta = (S - \sqrt{S^2-4})/2 > 0$ with $S = [2+(1-\rho^2)(K-1)]/\rho$.[45]

Proof See Appendix D. ∎

When $N = 1$ we have $\theta = \rho$ and (86) simplifies to $\Delta y_t^N = (1-\rho)\varepsilon_t$ implying that Δy_t^N is i.i.d.[46] As N grows, θ decreases, approaching 0 as N tends to infinity. When $N = \infty$, we recuperate Rotemberg's AR(1) process with first-order autocorrelation ρ.

As shown 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. This implies that the impuse response of Δy_t^N should be the same for $N = 1$ and $N = \infty$, which seems to contradict the particular cases of Proposition A.10 discussed above. What is going on? To answer this question, we take a brief detour to discuss the connection of Wold's representation with the missing persistence bias.

At stake is the fact that the ARMA representation in Proposition A.10 is a *linear representation* of the process followed by Δy_t^N, that is, it matches the first two moments (mean and covariances) of the actual process, but not necessarily higher moments. The impulse response functions implied by Proposition A.10 are not necessarily equal to the true impulse response, nor are the shocks implicit in this representation necessarily the shocks of economic interest.

The correct impulse response, that takes into account the non-linearities associated with lumpy adjustment, is quite different. To calculate this function, we consider a single unit and note that Δy_{t+k} is a response to Δy_t^* if and only if the first time the unit adjusts after the period t shock is in period $t+k$. It also follows from our Technical Assumptions that in this event the response is one-for-one. Thus

$$I_k = \Pr\{\xi_t = 0, \xi_{t+1} = 0, ..., \xi_{t+k-1} = 0, \xi_{t+k} = 1\} = (1-\rho)\rho^k. \tag{87}$$

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).

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 error. In the case we consider here we have $\mathrm{E}[\Delta y_t \Delta y_{t+1}] = 0$ and therefore $\Delta y_{t+1} - \mathrm{E}[\Delta y_{t+1}|\Delta y_t] = \Delta y_{t+1}$ so

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

[46] An alternative proof is obtained by applying the argument we used in Section 3.1 to show that the first-order autocorrelation is equal to zero to show that autocorrelations at higher lags are also zero.

that Wold's innovation at time $t+1$, Δy_{t+1}, differs from the innovation of economic interest, Δy^*_{t+1}. By contrast, from (92) and (93) we have

$$\mathrm{E}[\Delta y_{t+1}|x_t, \xi_t, x_{t-1}, \xi_{t-1}, ...] = (1-\rho)(1-\xi_{i,t})x_{it}$$

and it follows that the exact (non-linear) one-step ahead forecast error is equal to $(1-\rho)\Delta y^*_{i,t+1}$ and therefore linear in the shock of economic interest.

Wold's representation does not capture the entire process but only its first two moments.[47] 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 innovations in Wold's representation will not capture the response to the economic innovation of interest. This misidentification will be present in any VAR model including variables with lumpy adjustment.

We return to how to use Proposition A.10 to correct for the missing persistence bias. Using (86) to write Δy^N_t as an infinite moving average shows that its impulse response to ε-shocks satisfies:

$$I_k = \begin{cases} 1-\rho & \text{if } k=0 \\ \\ (1-\rho)(\rho-\theta)\rho^{k-1} & \text{if } k \geq 1. \end{cases} \tag{88}$$

Yet this is not the impulse response to the aggregate shock v^A_t, because ε_t in (86) is not v^A_t. As mentioned above, Wold's innovation is not the innovation of economic interest. The derivation of the true impulse response we did above 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 (86).

This suggests a straightforward approach to estimating the adjustment speed parameter, ρ: Estimate an ARMA(1,1) process (86) and read off the estimate of ρ (and the true impulse response) from the estimated AR-coefficient. That is, first estimate an ARMA model, next drop the MA polynomial and then make inferences about the implied dynamics using only the AR polynomial.

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

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

C.2 Instrumental Variables

Equation (86) in Proposition 1 suggests that lagged values of Δy and Δy^* (or components thereof) may be valid instruments to estimate ρ in a regression of the form

$$\Delta y^N_t = \text{const.} + \rho \Delta y^N_{t-1} + e_t.$$

[47] Of course, the first two moments determine the entire process if the process is Gaussian, the point here is that, with lumpy adjustment, the resulting aggregates are not Gaussian even if shocks are normal.

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

Proposition A11 (Instrumental Variables)

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

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

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

Proof See Appendix D. ∎

When applying this approach with actual data, the instruments turned out to be too weak and estimates were too imprecise.

D PROOFS OF PROPOSITIONS

In this appendix we present the proofs of propositions referred to in the main text and earlier in the appendix. We also include two sections with additional results referred to throughout the main text, one related to Rotemberg's Equivalence Result, the other to the expected response time index.

D.1 Proofs

We begin by proving a proposition, that includes Proposition 1 as the particular case where $\mu_A = 0$.

Proposition A12 (Aggregate Bias for Calvo Model With Drift)

Assume Technical Assumptions 1, 2 and 3 hold, where we allow for a non-zero mean μ_A *for aggregate shocks,* v^A*. Let* T *denote the time series length and* $\hat{\rho}$ *denote the OLS estimator of* ρ^N *in*

$$\Delta y_t^N = \text{const.} + \rho^N \Delta y_{t-1}^N + e_t. \tag{89}$$

Then, $\text{plim}_{T\to\infty}\hat{\rho}^N$ *depends on the weights* w_i *only through* N *and*

$$\text{plim}_{T\to\infty}\hat{\rho}^N = \frac{K}{1+K}\rho, \tag{90}$$

with

$$K \equiv \frac{\frac{1-\rho}{1+\rho}(N-1) - \left(\frac{\mu_A}{\sigma_A}\right)^2}{1 + \left(\frac{\sigma_I}{\sigma_A}\right)^2 + \frac{1+\rho}{1-\rho}\left(\frac{\mu_A}{\sigma_A}\right)^2}. \tag{91}$$

Proof The proof uses an auxiliary variable, x_{it}, equal to the unit's accumulated shocks since it last adjusted. It follows from the Technical Assumptions that x_{it} evolves according to:

$$x_{i,t+1} = (1-\xi_{it})x_{it} + \Delta y_{i,t+1}^*, \tag{92}$$

$$\Delta y_{it} = \xi_{it} x_{it}. \tag{93}$$

We first derive the following unconditional expectations:

$$\begin{aligned}
\mathrm{E}[x_{it}] &= \frac{\mu_A}{1-\rho}, &(94)\\
\mathrm{E}[\Delta y_{it}] &= \mu_A, &(95)\\
\mathrm{E}[\Delta y_t^N] &= \mu_A, &(96)\\
\mathrm{E}[x_{it}x_{jt}] &= \frac{1}{1-\rho^2}\left[\sigma_A^2+\frac{1+\rho}{1-\rho}\mu_A^2\right], &(97)\\
\mathrm{E}[x_{it}^2] &= \frac{1}{1-\rho}\left[\sigma_A^2+\sigma_I^2+\frac{1+\rho}{1-\rho}\mu_A^2\right], &(98)
\end{aligned}$$

where subindices i and j denote *different* units.

From (92) and the Technical Assumption in the main text we have:

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

The above expression leads to (94) once we note that stationarity of x_{it} implies $\mathrm{E}x_{i,t+1} = \mathrm{E}x_{it}$.

Equation (95) follows from (94) and Technical Assumption 3. Equation (96) follows directly from (95).

To derive (97), we note that, from (92)

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

where we used the Technical Assumptions, (94) and $i \neq j$. Noting that $x_{it}x_{jt}$ is stationary and therefore $\mathrm{E}[x_{it}x_{jt}] = \mathrm{E}[x_{i,t-1}x_{j,t-1}]$, the above expression leads to (97).

Finally, to prove (98), we note that, from (92) we have

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

where we used that $(1-\xi_{it})^2 = 1-\xi_{it}$, (94) 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 expressions for the four covariances involved in (12).

We have:

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

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

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

$$\begin{aligned} r_c(1) &= \text{Cov}(\Delta y_{i,t+1}, \Delta y_{jt}) = \text{E}[\xi_{i,t+1} x_{i,t+1} \xi_{jt} x_{jt}] - \mu_A^2 = (1-\rho)\text{E}[x_{i,t+1}\xi_{jt}x_{jt}] - \mu_A^2 \\ &= (1-\rho)\text{E}[\{(1-\xi_{it})x_{it} + \Delta y^*_{i,t+1}\}\xi_{jt}x_{jt}] - \mu_A^2 = \rho(1-\rho)^2\text{E}[x_{it}x_{jt}] + (1-\rho)\mu_A^2 - \mu_A^2 = \frac{1-\rho}{1+\rho}\rho\sigma_A^2. \\ r_c(0) &= \text{Cov}(\Delta y_{it}, \Delta y_{jt}) = \text{E}[\xi_{it}x_{it}\xi_{jt}x_{jt}] - \mu_A^2 = (1-\rho)^2\text{E}[x_{it}x_{jt}] - \mu_A^2 = \frac{1-\rho}{1+\rho}\sigma_A^2. \end{aligned}$$

Finally, the variance term is obtained as follows:

$$r_a(0) = \text{Var}(\Delta y_{it}) = \text{E}[\xi_{it}^2 x_{it}^2] - \mu_A^2 = \text{E}[\xi_{it}x_{it}^2] - \mu_A^2 = (1-\rho)\text{E}[x_{it}^2] - \mu_A^2 = \sigma_A^2 + \sigma_I^2 + \frac{2\rho}{1-\rho}\mu_A^2.$$

Substituting the above expressions for $r_a(1)$, $r_a(0)$, $r_c(1)$ and $r_c(0)$ in (12) leads to (90) and (91) and completes the proof. ∎

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:

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

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 (92), yielding $\sigma_A^2 + \sigma_I^2$ and σ_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 four covariances we calculated above, when proving Proposition 1. We have all the terms for the expression above for $\hat{b}_1$, the remainder of the proof is some tedious but straightforward algebra. ∎

Proof of Proposition A.10

To prove that Δy_t^N follows an ARMA(1,1) process with autoregressive coefficient ρ, it suffices to show that the process' autocorrelation function, γ_k, satisfies:[48]

$$\gamma_k = \rho\gamma_{k-1}, \qquad k \geq 2. \tag{99}$$

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

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

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

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

To show that (99) holds, we note that:

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

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

where in the fourth step we assumed $k \geq 2$, since we used that $\xi_{i,t+k-1}$ and ξ_{jt} are independent even when $i = j$. Noting that $\gamma_k = (\mathrm{E}[\Delta y^N_{t+k}\Delta y^N_t]-\mu_A^2)/\mathrm{Var}(\Delta y_t)$ and using the above identity yields (99) and concludes the proof. ∎

Proof of Proposition A.11

We have:

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

Similarly

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

Subtracting the latter from the former and rearranging terms yields

$$\Delta y^N_t = \rho\Delta y^N_{t-1}+(1-\rho)\Delta y^{*N}_t+\epsilon^N_t \tag{100}$$

with

$$\epsilon^N_t = \sum_i w_i\left[(\xi_{it}-1+\rho)(y^*_{it}-y_{i,t-1})-(\xi_{i,t-1}-1+\rho)(y^*_{i,t-1}-y_{i,t-2})\right]. \tag{101}$$

The extra term ϵ^N_t on the r.h.s. of (100) explains why Δy^N_{t-1} is not a valid instrument: Δy^N_{t-1} is correlated with ϵ^N_t because both include $\xi_{i,t-1}$ terms. Of course, ϵ^N_t tends to zero as N tends to infinity: its mean is zero and a calculation using many of the expressions derived in the proof of Proposition 1 shows that

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

It follows from (100), (101) and Technical Assumption 3 that ϵ_t is uncorrelated with Δy^*_s, for all s, which implies that Δy^*_{t-s} is a valid instrument for $s \geq 1$. And since $\Delta y_{i,t-k}$ are uncorrelated with ξ_{it} and $\xi_{i,t-1}$ for $k \geq 2$, we have that lagged values of Δy, with at least two lags, are valid instruments as well. ∎

Proof of Proposition 9

The equation we estimate is:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t, \tag{102}$$

while the true relation is that described by (80) and (81).

It is easy to see that the second term on the right hand side of (80) denoted by w_t in what follows, is uncorrelated with Δy_{t-k}, $k \geq 1$. It follows that estimating (102) is equivalent to estimating (80) 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.2 Rotemberg's Equivalence Result

In this appendix we state Rotemberg's Equivalence Result, which we used to in Section 2. We also derive expressions that follow from this result and that are used later in this appendix.

Proposition A13 (Rotemberg's Equivalence Result)

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

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

where β denotes the agent's discount factor and y^*_t denotes an exogenous process.[49] *We then have*

$$\tilde{y}_t = (1-\beta\rho) \sum_{k\geq 0} (\beta\rho)^k \mathrm{E}_t y^*_{t+k}. \tag{103}$$

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

$$y^\infty_t = \rho y^\infty_{t-1} + (1-\rho)\tilde{y}_t. \tag{104}$$

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

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

where $c > 0$ captures the relative importance of quadratic adjustment costs. We then have that there

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

exists $\rho' \in (0,1)$ *and* $\delta \in (0,1)$ *s.t.*[50]

$$y_t^\infty = \rho' y_{t-1}^\infty + (1-\rho')\hat{y}_t, \tag{105}$$

with

$$\hat{y}_t = (1-\delta)\sum_{k\geq 0} \delta^k \mathrm{E}_t y_{t+k}^*. \tag{106}$$

Finally, and this is Rotemberg's equivalence result and contribution, a comparison of (103)-(104) *and* (105)-(106) *shows that an econometrician working with aggregate data cannot distinguish between the Calvo model and the Quadratic Adjustment Costs model described above:* ρ' *plays the role of* ρ *and* δ *the role of* $\beta\rho$.

Proof See Rotemberg (1987). ∎

What we use in this paper from Rotemberg's result is not the aggregate equivalence between lumpy adjustment and quadratic adjustment costs. What we use is that in a model with Calvo adjustment and an infinite number of agents, the aggregate of interest is equal to a distributed lag of aggregate shocks and therefore a linear function of these shocks.

Corollary A2 *Under the assumptions of the Calvo Model in Proposition 13.*

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

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

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

$$y_t^\infty = (\rho+\psi) y_{t-1}^\infty - \rho\psi y_{t-2}^\infty + \frac{(1-\rho)(1-\beta\rho)}{1-\beta\rho\psi} e_t. \tag{107}$$

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

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

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

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

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

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

Proof Straightforward. ∎

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

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

D.3 Expected Response Time

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

$$\tau \;\equiv\; \frac{\sum_{k\geq 0} k I_k}{\sum_{k\geq 0} I_k}, \tag{108}$$

with

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

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

Lemma A2 (τ for an Infinite MA) *Consider a second order stationary stochastic process*

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

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

Then:

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

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

Proposition A14 (τ for an ARMA Process) *Assume Δy_t follows an ARMA(p,q):*

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

where $\Phi(z) \equiv 1 - \sum_{k=1}^{p}\phi_k z^k$ and $\Theta(z) \equiv 1 - \sum_{k=1}^{q}\theta_k z^k$ have all their roots outside the unit disk. The assumptions regarding the ϵ_t's are the same as in Lemma A2.

Define τ as in (108)*. 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 A2 with $\Theta(z)/\Phi(z)$ in the role of $\Psi(z)$ we then have:

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

Proposition A15 (τ for a Lumpy Adjustment Process) *Consider Δy_t in the standard Calvo lumpy adjustment model* (5) *and τ defined in* (108). *Then $\tau = \rho/(1-\rho)$.*

Proof $\partial\Delta y_{t+k}/\partial\Delta y_t^*$ is equal to one when the unit adjusts at time $t+k$, not having adjusted between times t and $t+k-1$, and is equal to zero otherwise. Also, from (88) we have that

$$I_k \equiv \mathrm{E}_t\left[\frac{\partial\Delta y_{t+k}}{\partial\Delta y_t^*}\right] = (1-\rho)\rho^k.$$

The expression for τ now follows easily. ∎

E QUANTITATIVE MODEL

E.1 An *Ss* and Calvo Model

This section gives the details of the baseline menu cost and Calvo models we use in Sections 3, 4 and 6. This model is a single sector version of the GE Ss/Calvo model in Nakamura and Steinsson (2010).

E.1.1 Households

The household side of the model is straightforward:

$$\max_{n_t, c_{it}} E_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - \omega n_t\right],$$

subject to

$$\int_0^1 p_{it} c_{it} di \le W_t n_t + \int_0^1 \Pi_{it},$$

where

$$C_t = \left(\int_0^1 (c_{it})^{\frac{\theta-1}{\theta}} di\right)^{\frac{\theta}{\theta-1}}$$

is a Dixit-Stiglitz aggregator of consumption goods c_{it}, p_{it} is the price of good i, n_t is the household's labor supply, ω is the disutility of labor, W_t is the nominal aggregate wage, Π_{it} is the profits the household receives from owning firm i, and θ is the elasticity of substitution.

Given firm prices, household demand is given by:

$$c_{it} = \left(\frac{p_{it}}{P_t}\right)^{-\theta} C_t,$$

where P_t is the Dixit-Stiglitz price index:

$$P_t = \left(\int_0^1 (p_{it})^{1-\theta} di\right)^{\frac{1}{1-\theta}}.$$

The first order condition for labor supply gives:

$$\omega = \lambda_t W_t$$

where λ is the multiplier on the budget constraint. The consumption first order condition implies that

$$c_{it}^{-1/\theta}\left(\int_0^1 c_{it}^{(\theta-1)/\theta}\,\mathrm{d}i\right)^{-1} = \lambda p_{it}.$$

Going through a bit more algebra, we get that $\lambda_t = 1/C_t P_t$ so the real wage is given by $W_t/P_t = \omega C_t$.

E.1.2 Firms

Turning to the firms' problem, they produce using a linear production function in labor

$$y_{it} = z_{it} l_t^i,$$

where firm i's idiosyncratic productivity z_{it} evolves according to

$$\log z_{it} = \rho_z \log z_{it-1} + \varepsilon_{it}; \quad \varepsilon_{it} \sim N(0, \sigma_z^2).$$

Firms pay a fixed menu cost f in units of labor in order to adjust their nominal price. Given these constraints, firm i's problem is to choose prices to maximize discounted profits.

$$\max_{p_{it}} E_t \sum_{t=0}^{\infty} Q^t \Pi_{it},$$

where $Q = \beta U'(C')/U'(C) = \beta C/C'$ and flow firm profits are given by:

$$\Pi_t^i = \left(\underbrace{\frac{p_{it}}{P_t}}_{\text{Unit Revenues}} - \underbrace{\frac{W_t}{z_{it} P_t}}_{\text{Unit Cost}}\right)\underbrace{\left(\frac{p_{it}}{P_t}\right)^{-\theta} C_t}_{\text{Demand}} - \underbrace{\kappa \frac{W_t}{P_t} I_{p_{it} \neq p_{it-1}}}_{\text{Menu Cost if Adjusting}}.$$

E.1.3 Equilibrium and Laws of Motion

Nominal Demand is assumed to be a random walk in logs:

$$\log S_{t+1} = \log S_t + \mu + \varepsilon_t,$$

with the ε_t i.i.d. zero mean, normal, independent from the idiosyncratic shocks defined above. The aggregate price level will be a function of aggregate spending and the initial distribution of firms $P_t = \varphi\left(\chi\left(p_{-1}, z\right), S\right)$. Given the density of firms χ, φ, and the evolution of χ, we write down the firm problem as:

$$
\begin{aligned}
V(p_{-1}, z; \chi(p_{-1}, z), S) &= \max\{V^a(z; \chi(p_{-1}, z), S), V^n(p_{-1}, z; \chi(p_{-1}, z), S)\} \\
V^a(z; \chi(p_{-1}, z), S) &= \max_p \left(\frac{p}{P} - \frac{\omega \frac{S}{P}}{z}\right)\left(\frac{p}{P}\right)^{-\theta} \frac{S}{P} - \kappa\omega\frac{S}{P} \\
&\quad + \beta E_{z,S'} \frac{\frac{S}{P}}{\frac{S'}{P'}} V\left(\frac{p}{P'}, \rho_z z + \varepsilon; \chi'(p'_{-1}, z'), S'\right) \\
V^n(p_{-1}, z; \chi(p_{-1}, z), S) &= \left(\frac{p_{-1}}{P} - \frac{\omega \frac{S}{P}}{z}\right)\left(\frac{p_{-1}}{P}\right)^{-\theta} \frac{S}{P} \\
&\quad + \beta E_{z,S'} \frac{\frac{S}{P}}{\frac{S'}{P'}} V\left(\frac{p_{-1}}{P}, \rho_z z + \varepsilon; \chi'(p'_{-1}, z'), S'\right) \\
\text{with } P &= \varphi(\chi(p_{-1}, z), S) \;\&\; \chi'(p'_{-1}, z') = \Gamma(\chi(p_{-1}, z), S').
\end{aligned}
$$

In order to make this problem tractable, we follow Krusell-Smith (1998) and assume that we can accurately forecast how the aggregate price level evolves using the simple log-linear equation:

$$
\log\frac{P}{S} = \gamma_0 + \gamma_1 \log\frac{P_{-1}}{S}.
$$

Consistent with Nakamura and Steinsson (2010), we find that this update rule works well in practice and delivers values for R^2 above 99%.

E.1.4 Models Nested in this Framework

As pointed out by Nakamura and Steinsson (2010), a nice feature of this model is that by making different assumptions about κ, the model nests the standard menu cost and Calvo models, as well as models that are quantitatively similar to multi-product menu cost models such as Midrigan (2011) and Alvarez, Le Bihan and Lippi (2016). In particular:

- *Standard menu cost model*: constant κ
- *Standard Calvo model*: with probability ρ, $\kappa = \infty$, with probability $1-\rho$, $\kappa = 0$, where $1-\rho$ is the frequency of price adjustment
- *CalvoPlus model*: with probability ρ, $\kappa = M_{\text{large}}$, with probability $1-\rho$, $\kappa = M_{\text{small}}$, where M_{large} and M_{small} are large and small numbers respectively.
- *Multi-product menu cost model*: κ is i.i.d. and drawn from some distribution, $G(.)$. [51]

E.1.5 Structural Interpretation of the Inflation Equation

In the Calvo version of our model where both the nominal shock, S_t, and the idiosyncratic shock, z_{it}, follow a random walk and steady state inflation is zero[52], we can provide a structural interpretation of the reduced form equations introduced in Section 2 and derive an explicit expression of our estimating equation (7) from the main text.

[51] In this case, we have to slightly modify the firms problem below so that firms take expectations of firm value functions with respect to G. For simplicity we did not add this detail since it is not relevant for the other three models.

[52] This assumption can easily be relaxed but it makes the formulas more cumbersome.

Firms choose P_{it}^{*} to maximize the following discounted sum:

$$\max_{P_{i,t}^{*}} \sum_{k=0}^{\infty} (\beta\rho)^{k} \left[\left(\frac{C_t}{C_{t+k}} \right) \left(P_{it}^{*(1-\theta)} P_{t+k}^{\theta-1} C_{t+k} - \frac{W_{t+k}}{z_{it+k}} P_{it}^{*-\theta} P_{t+k}^{\theta-1} C_{t+k} \right) \right].$$

which simplifies to:

$$\max_{P_{i,t}^{*}} \sum_{k=0}^{\infty} (\beta\rho)^{k} \left[P_{t+k}^{\theta-1} \left(P_{it}^{*(1-\theta)} - \frac{W_{t+k}}{z_{it+k}} P_{it}^{*-\theta} \right) \right] C_t$$

The equation for the optimal reset price is:

$$P_{it}^{*} = \frac{\sum_{k=0}^{\infty} (\beta\rho)^{k} E_t P_{t+k}^{\theta-1} \frac{W_{t+k}}{z_{it+k}}}{\sum_{k=0}^{\infty} (\beta\rho)^{k} P_{t+k}^{\theta-1}}.$$

If there are no pricing frictions the firm sets its optimal price as a static markup over marginal costs. With pricing frictions, the optimal is a function of expected future nominal costs and some aggregate factors. Taking logs and a first order approximation in differences to eliminate constants gives:

$$p_{it}^{*} = (1-\rho\beta) \sum_{k=0}^{\infty} (\beta\rho)^{k} E_t \left[w_{t+k} - \log z_{it+k} \right] \tag{109}$$

Using the fact that the nominal wage $W_t = \omega S_t$ and substituting in the above equation:

$$p_{it}^{*} = (1-\rho\beta) \sum_{k=0}^{\infty} (\beta\rho)^{k} E_t \left(s_{t+k} - \log z_{it+k} \right).$$

The target price therefore is

$$p_{it}^{*} = (1-\rho\beta) \sum_{k=0}^{\infty} (\beta\rho)^{k} E_t \left[s_{t+k} - \log z_{it} \right] = s_t - \log z_{it}.$$

Note that the average chosen price level is

$$\int p_{i,t}^{*} di = s_t.$$

Since firms choose their prices optimally, when firms adjust, their adjustment will be equal to the sum of past sequence of shocks since they last adjusted. Otherwise, their change in prices will be zero. To see this, consider a firm that adjusts in period t but last adjusted k periods ago. Then they adjust by the following amount:

$$\begin{aligned} p_{it}^{*} - p_{it-k}^{*} &= (p_{it}^{*} - p_{it-1}^{*}) + (p_{it-1}^{*} - p_{it-2}^{*}) + \ldots + (p_{it-k+1}^{*} - p_{it-k}^{*}) \\ &= (\varepsilon_t - \varepsilon_{it}) + (\varepsilon_{t-1} - \varepsilon_{it-1}) + \ldots + (\varepsilon_{t-k+1} - \varepsilon_{it-k+1}) \end{aligned}$$

This provides a structural interpretation of the reduced form equations introduced in Section 2 (see page 4 of the main text).

Finally, we can aggregate to get an expression for aggregate inflation. A log-linear approximation for the aggregate price level is given by the following expression

$$p_{t+1} = \rho p_t + \left(1-\rho\right)\int p^*_{it} di = \rho p_t + \left(1-\rho\right) s_t.$$

Then the inflation satisfies

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

since the idiosyncratic shocks have a zero mean. This provides a micro-foundation for our main estimating equation (7). Interestingly, a similar equation holds in the Ss model of Gertler and Leahy (2008). The main difference is that the coefficient on lagged inflation is no longer the frequency of non-adjustment.[53]

E.2 Calibration Details

E.2.1 Single Sector Models

The details of our single-sector Calvo and Ss calibration are as follows. We broadly follow the literature in the moments we target, with one exception. In addition to the traditionally targeted moments, we also include the sampling error among the moments to be matched in all of our calibration exercises.[54] We did this for two reasons. First, the forces that lead to large sampling errors (low number of observations, large idiosyncratic shocks) are among the factors that lead to a large missing persistence bias.[55] Second, Proposition 4 in the main text identifies the sampling error as a key moment to gauge the size of the missing persistence bias. This proposition provides an estimate for the (relative) bias as a simple expression that involves only two moments: the standard deviation of the aggregate inflation series and the sampling error. We computed our estimates of sampling error using a simple bootstrap procedure, which is consistent with how the BLS computes the sampling error for the CPI.

We target the following 6 moments in all of our calibrations (both Calvo and Ss):

Targeted Moments

1. Frequency of adjustment: 0.14 (source: CPI micro data,1988:2-2007:12)
2. Sampling error of one month inflation: 0.00040 (Source: https://www.bls.gov/cpi/tables/variance-estimates/home.htm – we use the estimate for the entire CPI)
3. Average size of price increases: 8.9% (Source: Klenow and Krystov, 2010)
4. Average size of price decreases: 9.4% (Source: Klenow and Krystov, 2010)
5. Fraction of price changes that are price increases: 0.57 (Source: Klenow and Krystov, 2010)
6. Standard deviation of inflation rate: 0.0022 (Source: CPI-U, 1988:2-2007:12)

[53] See equation 3.24 of the NBER working paper version of their paper: http://www.nber.org/papers/w11971.pdf.

[54] We thank one of referees for this suggestion.

[55] Notice though that the biases induced by sampling error are distinct from the missing persistence bias because the missing persistence bias would be zero if adjustment were frictionless even with a finite numbers of observations, while sampling error would not.

We fix a number of other parameters because they are either directly observable or they have a direct correspondence to an observable. These are:

Externally Calibrated Parameters

1. The mean of the aggregate shock, μ_A. We set this equal to 0.002 which is the mean of monthly CPI inflation.[56]

2. Standard deviation of aggregate shock, σ_A. Following Nakamura and Steinsson (2008) and Vavra (2014), we match this moment to the standard deviation of nominal GDP, which is the analog of the aggregate monetary shock in the model. We set this equal to 0.0037.

3. The frequency of adjustment, ρ, in Calvo model calibrations.

4. The effective number of observations, N. We often allow this parameter to vary depending on the purpose of the calibration, but if we are doing simulations that are representative of the entire U.S. CPI we use $N = 15,000$ because this is close to the average effective number of active price quotes per month in the CPI micro data.

We used these moments to calibrate the following "internal" parameters that are different depending on the model:

Parameters to be Calibrated

1. Calvo model: the persistence of idiosyncratic shock, ρ_I (in the random walk case this is fixed and equal to one) and the standard deviation of the idiosyncratic shock, σ_I.

2. GE *Ss* model: the menu cost, κ, the persistence of idiosyncratic shock, ρ_I (in random walk case this is fixed) and the standard deviation of the idiosyncratic shock, σ_I.

3. PE *Ss* models: the location of the adjustment bands (S and s), the persistence of idiosyncratic shock, ρ_I (in the random walk case this is fixed) and the standard deviation of the idiosyncratic shock, σ_I.

We consider two objective functions for weighting these moments:

Weights

1. Equal weights on all moments

2. Triple the weight on the frequency and the sampling error

Our results were not sensitive to using other reasonable weighting schemes.

[56] This is justified by a straightforward extension of Proposition A.6 to the case with non-zero mean for aggregate shocks.

E.2.2 Multi-Sector Models

The details of our multi-sector Calvo and Ss models calibration used in Section 6 are as follows. We calibrate a 66 sector version of each 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 same six moments that we match in the single sector versions of the model: the frequency of adjustment, the average size of price increases and of price decreases, the fraction of price changes that are price increases, the standard deviation of the sectoral inflation rate and sampling error of one-month sectoral inflation. In the model, the effective number of firms in each sector is given by the median (across time) effective number of firms for that sector in the micro BLS data and each firm was simulated for 238 periods, which is the number of periods in the underlying data.

Table 7 shows basic descriptive statistics for the simulated model. The reported statistics are medians across the 66 sectors, suggesting that both models do a good job matching moments across sectors.

Table 7: CALIBRATION DETAILS: MULTI-SECTOR CALVO AND SS

Calibration Results: Basic Statistics

	CPI	Calvo	Ss
Frequency of monthly adjustment	0.068	0.066	0.067
Fraction of price changes that are positive	0.669	0.583	0.694
Average size of price increases	7.997	8.362	8.609
Average size of price decreases	9.073	6.624	8.516
Std deviation of sectoral inflation	0.005	0.002	0.004
Sampling error for one month inflation	0.310	0.316	0.385

F CALIBRATION RESULTS

F.1 Overview

In this appendix we assess the magnitude of the missing persistence bias for the US CPI. We focus on inflation because this is the variable of interest in the two applications in Section 6. We assess this bias for the Calvo model from Section 3, the small σ_A version of the Ss model from Section 4 and standard Ss models. Both general and partial equilibrium models are considered.

Table 8 summarizes this bias estimates for five representative models of the 14 models we calibrated. The upper half of the table reports average values for the estimates of $\hat{\rho}^N$ for different values of the effective number of units, N. These values should be compared with those reported in the last column, which corresponds to the value obtained with an infinite number of units.[57] The lower half of Table 8 reports the relative bias, $(\hat{\rho}^N - \rho^\infty)/\rho^\infty$, where ρ^∞ denotes the (theoretical) first-order

[57] In some cases $N = \infty$ corresponds to simulations with $N = 40{,}000$. The bias estimates we obtained will therefore underestimate the true bias in these cases, even though the difference is likely to be very small.

correlation for an aggregate with an infinite number of units (ρ_c in the notation of Proposition 3). This measure for the magnitude of this bias is conservative, working with alternative measures such as the log-difference of the half-life of a shock, leads to larger values for this bias.

The first model in Table 8 is a Calvo model calibration that follows our baseline assumptions (Δy^* is i.i.d.). The second model is the small σ_A *Ss* model. The third model is the standard *Ss* counterpart to our Calvo model. The fourth and fifth models consider PE and GE versions of the same *Ss* model where the idiosyncratic shock is allowed to follow an AR(1), a standard assumption in the literature. In the last three models, the probability of adjusting depends on the history of idiosyncratic and aggregate shocks, by contrast with the second model where this probability only depends on idiosyncratic shocks.[58]

Table 8: CALIBRATION RESULTS: SUMMARY

Measure	Calibration	Effective number of agents (N)					
		100	400	1,000	4,000	15,000	∞
$\hat{\rho}^N$	*CPI Data (bootstrap)*	*0.051*	*0.127*	*0.200*	*0.289*	*0.316*	0.330
	Calvo Random Walk	0.022	0.083	0.182	0.447	0.690	0.860
	Ss Small σ_A	0.033	0.131	0.270	0.559	0.708	0.862
	Ss Random walk	0.139	0.286	0.359	0.410	0.422	0.431
	Ss AR(1) PE	0.103	0.258	0.342	0.405	0.421	0.432
	Ss AR(1) GE	0.064	0.218	0.304	0.388	0.421	0.435
Relative Error $(\hat{\rho}^N - \rho^\infty)/\rho^\infty$	*CPI Data (bootstrap)*	*-0.845*	*-0.615*	*-0.394*	*-0.124*	-0.042	0.000
	Calvo Random Walk	-0.975	-0.904	-0.788	-0.481	-0.198	0.000
	Ss Small σ_A	-0.961	-0.848	-0.687	-0.352	-0.178	0.000
	Ss Random Walk	-0.677	-0.336	-0.167	-0.048	-0.020	0.000
	Ss AR(1) PE	-0.761	-0.402	-0.207	-0.060	-0.025	0.000
	Ss AR(1) GE	-0.853	-0.499	-0.301	-0.108	-0.032	0.000

The mean, median and maximum number of effective observations in the 66 CPI sectors we consider in Section 6 are 187, 142 and 980 respectively.[59] This bias is large at these levels of aggregation for all models considered, this is the main message of this section. In particular, for $N = 100$ the relative bias is above 60% in all cases, and for $N = 400$ it is larger than 30% for all models. By contrast, at the aggregate CPI level ($N = 15,000$), this bias becomes close to negligible for standard Ss models while it remains significant (around 20%) for the Calvo model and the small σ_A *Ss* model.[60] The conclusion, then, is that this bias matters for most 2-digit sectors and could potentially be relevant at higher levels of aggregation.

The first row in Table 8 reports estimates for $\hat{\rho}^N$ obtained from the actual CPI micro database via bootstrap simulations.[61] Consistent with the prediction of the missing persistence bias, these

[58] All models reported in this summary table consider equal weights on all moments, with the exception of the second model, where we consider the calibration that gives more weight to the fraction of units adjusting. We do this to facilitate comparison across models, since in this case the calibration with equal weights has a fraction of adjusters that is significantly larger than in the remaining models (0.216 vs. an average of 0.14).

[59] Our definition of sectors is close to a two digit level of disaggregation.

[60] For the half-life of shocks the relative bias for the last two models slightly above 50%.

[61] Specifically, we randomly sample N price change observations in each month (including zeros) and use this sub-

estimates increase with the level of aggregation, from 0.051 when $N = 100$ to 0.316 when $N = 15,000$, which corresponds to the effective number of prices used when calculating the CPI. The following rows report results for five models.

The regression coefficient $\hat{\rho}^N$ for actual inflation is not among the moments considered in the calibration exercises and therefore provides a useful benchmark to compare models. The upper half of Table 8 shows that the first two models match the regression coefficients best when $N = 400$ and $N = 1,000$. By contrast, the remaining three models do a better job when $N = 4,000$ and $N = 15,000$. Overall, this suggests that Ss models do much better than Calvo models at matching this moment of the data.

Comparisons across models also provide some interesting insights. The Calvo Random Walk and the small σ_A Ss model, which also assumes a random walk for y^*, lead to similar bias estimates. By contrast, this bias is much smaller for the standard Ss model with a random walk. This may be due to the fact that, as noted in Appendix A, the response of inflation to an aggregate shock upon impact will be much larger under standard Ss models than under the Calvo or the small σ_A Ss model. To the extent that the impulse response decays approximately at a geometric rate, as is the case for the models calibrated for US CPI in this paper, this will imply a considerably larger value for ρ^∞ for standard Ss models. Also, the Ss AR(1) models both generate similar relative error estimates suggesting that GE does not have a first order effect on the magnitude of this bias.

F.2 Bias with Calvo Adjustment

In this section we discuss the relevance of our theory in the price-setting context when the true data generating process is a Calvo model. If in addition we assume that each firms idiosyncratic shock follows a random walk, then Proposition 1 and its extensions in Appendix D to the case with non-zero drift (Proposition A.12) applies.

We consider four possible calibration of the Calvo model:

1. AR(1): ($1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 0.85$, $\sigma_I = 0.1088$)
2. AR(1) (extra weight): ($1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 0.85$, $\sigma_I = 0.12$)
3. Random walk: ($1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.0616$)
4. Random walk (extra weight): ($1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.0661$)

Table 9 reports how the estimated ρ varies with the effective number of units, N, for different calibrations. As in the summary table, the first half shows the implied $\hat{\rho}^N$ for each calibration for different values of N. The second half reports the relative bias, $(\hat{\rho}^N - \rho^\infty)/\rho^\infty$, for each value of N. A larger (more negative) value means a bigger bias.

The table shows that the for reasonable parameterizations of the Calvo model, the missing persistence bias is large. The median relative error across all four calibrations is 96% for $N = 100$, 85% for $N = 400$ and 70% for $N = 1,000$. When $N = 15,000$, which is the average effective number of active price quotes per month in the entire CPI, the median bias is 14%.

sample to compute a time-series of inflation rates, $\hat{\pi}_t$ making sure that the implied frequency of adjustment is similar across samples. We then estimate an AR(1) on this inflation series as a measure of persistence. We repeat this process 500 times and display the mean estimate and standard error.

Table 9: CALVO MODEL CALIBRATION RESULTS

Measure	Calibration	Effective number of agents (N) 100	400	1,000	4,000	15,000	∞
$\hat{\rho}^N$	AR(1)	0.065	0.220	0.408	0.671	0.798	0.853
	AR(1) (extra weight)	0.049	0.173	0.342	0.618	0.783	0.856
	Random Walk	0.022	0.083	0.182	0.447	0.690	0.860
	Random Walk (extra weight)	0.014	0.056	0.128	0.355	0.623	0.860
Relative	AR(1)	-0.924	-0.742	-0.522	-0.213	-0.065	0.000
Error:	AR(1) (extra weight)	-0.943	-0.798	-0.601	-0.278	-0.085	0.000
$(\hat{\rho}^N - \rho^\infty)/\rho^\infty$	Random Walk	-0.975	-0.904	-0.788	-0.481	-0.198	0.000
	Random Walk (extra weight)	-0.984	-0.935	-0.851	-0.588	-0.275	0.000
Median Relative Error		-0.959	-0.851	-0.695	-0.380	-0.142	0.000

F.3 Bias With State-Dependent Adjustment

In this section we discuss the relevance of our theory in the price-setting context when the true data generating process is a menu-cost or Ss model. We consider two versions of this model. The first version, which we refer to as "small σ_A" Ss model, assumes firms only adjust in response to idiosyncratic shocks. As shown in Appendix A1, under these conditions, the only difference between this bias for the Calvo and Ss models comes from differences in the distributions of the number of periods since agents last adjusted (the γ_k in Proposition A.3). The second model allows the probability of adjustment to depend on both aggregate and idiosyncratic shocks, both in PE and in GE versions.

We consider two possible calibrations of the small σ_A model. We match the same six moments as before. The extra weight calibration indicate more weight (3 times) for the frequency of adjustment and sampling error moments.

1. Small σ_A Ss: s = -0.069, $S = 0.064$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\sigma_I = 0.044$, $\rho_I = 1.0$.

2. Small σ_A Ss (extra weight): s = -0.103, $S = 0.095$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\sigma_I = 0.048$, $\rho_I = 1.0$.

The first half of Table 10 reports how the estimated $\hat{\rho}^N$ varies with the effective number of units, N. The first two rows show the implied $\hat{\rho}$, the following two rows report the relative bias, $(\hat{\rho} - \rho)/\rho$, for each value of N.

Table 10 shows that this bias is smaller for the small σ_A Ss model than for the Calvo model. For example, for $N = 1,000$ the former have a median relative bias of 61% compared with 70% for the latter. As we will see shortly, these differences are much smaller than the ones we find when comparing standard Ss models with the Calvo model.

Next we consider 8 calibrations for the second type of Ss model. The first 6 use the GE Ss model described in Appendix E.1 while the last 2 target the Ss bands directly. We match the same six observed moments that we mathched before. The extra weight calibrations indicate more weight (3 times) was given to the frequency of adjustment and sampling error moments.

Table 10: *Ss* MODELS (SMALL σ_A ASSUMPTION)

Measure	Calibration	Effective number of agents (N)					
		100	400	1,000	4,000	15,000	∞
$\hat{\rho}^N$	Ss	0.061	0.207	0.374	0.622	0.716	0.796
	Ss (extra weight)	0.033	0.131	0.270	0.559	0.708	0.862
Relative error	Ss	-0.923	-0.740	-0.530	-0.219	-0.101	0.000
$(\hat{\rho}^N - \rho^\infty)/\rho^\infty$	Ss (extra weight)	-0.961	-0.848	-0.687	-0.352	-0.178	0.000
Median Relative Error		-0.942	-0.794	-0.609	-0.286	-0.140	0/000

Table 11: *Ss* MODELS: GENERAL CASE

Measure	Calibration	Effective number of agents (N)					
		100	400	1,000	4,000	15,000	∞
$\hat{\rho}$	AR(1) PE	0.103	0.258	0.342	0.405	0.421	0.432
	AR(1) GE	0.064	0.218	0.304	0.388	0.421	0.435
	AR(1) (extra weight)	0.025	0.184	0.306	0.423	0.456	0.480
	NS 2010	0.166	0.376	0.480	0.553	0.570	0.583
	Random Walk	0.139	0.286	0.359	0.410	0.422	0.431
	Random Walk (extra weight)	0.089	0.228	0.321	0.402	0.424	0.439
	Ss PE	0.118	0.260	0.339	0.399	0.413	0.423
	Ss PE (extra weight)	0.077	0.205	0.299	0.386	0.410	0.428
Relative	AR(1) PE	-0.761	-0.402	-0.207	-0.060	-0.025	0.000
Error:	AR(1) GE	-0.853	-0.499	-0.301	-0.108	-0.032	0.000
$(\hat{\rho} - \rho)/\rho$	AR(1) (extra weight)	-0.947	-0.616	-0.362	-0.119	-0.051	0.000
	NS 2010	-0.714	-0.355	-0.177	-0.050	-0.021	0.000
	Random Walk	-0.677	-0.336	-0.167	-0.048	-0.020	0.000
	Random Walk (extra weight)	-0.796	-0.481	-0.268	-0.084	-0.035	0.000
	Symmetric Ss	-0.721	-0.385	-0.199	-0.057	-0.024	0.000
	Symmetric Ss (extra weight)	-0.820	-0.521	-0.301	-0.098	-0.042	0.000
Median Relative Error		-0.779	-0.442	-0.238	-0.072	-0.024	0.000

Table 11 reports the results. The first eight rows show the implied $\hat{\rho}$, for each calibration for different values of N. We obtain this bias by simulating the model. The following eight rows report the relative bias, $(\hat{\rho}^N - \rho)\rho$, for each value of N. A larger (more negative) value means a bigger bias. The table shows that the for reasonable parameterizations of these standard Ss models, the missing persistence bias is large, though smaller than in the previous two models. While this bias is sizable at the sectoral level (the median relative biases across models for $N = 100$, $N = 400$ and $N = 1,000$ are 78%, 44% and 24%), this bias is negligible when $N = 15,000$. Thus the small σ_A assumption that is frequently made in the literature is not innocuous in this context.

1. AR(1) : ($\kappa = 0.02$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 0.80$, $\sigma_I = 0.0443$)
2. AR(1) GE: ($\kappa = 0.02$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 0.80$, $\sigma_I = 0.0443$)
3. AR(1) (extra weight): ($\kappa = 0.040$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 0.65$, $\sigma_I = 0.0672$)
4. NS 2010: ($\kappa = 0.0245$, $\mu_A = 0.0021$, $\sigma_A = 0.0037$, $\rho_I = 0.66$, $\sigma_I = 0.0425$)
5. Random walk: ($\kappa = 0.0214$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.036$)
6. Random walk (extra weight): ($\kappa = 0.0443$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.0516$)
7. Ss PE: s = -0.074, $S = 0.069$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\sigma_I = 0.035$, $\rho_I = 1.0$.
8. Ss PE (extra weight): s = -0.099, $S = 0.090$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\sigma_I = 0.046$, $\rho_I = 1.0$.

F.4 The Relevance of the Missing Persistence Bias for Other Variables

In this section, we consider two other macroeconomic variables where lumpy microeconomic adjustment has been well established —employment and investment— and use Proposition 4 to provide estimates of the magnitude of the missing persistence bias for each of these variables.

For employment, we use quarterly CES data from the BLS. This data is available for the 1990q1-2016q1 period. For every series we compute the standard deviation of change in log employment.[62] Estimates for the sampling error for each sector come directly from the BLS.[63] We make one adjustment to these published sampling errors. If you compare the relative sampling errors to their respective errors in levels, they imply a benchmark level of employment that is too low. For example, they imply that the aggregate level of employment is 35 million not 140 million. We thus adjust our sampling errors downward so that we hit the actual employment benchmarks. Since we are using smaller sampling errors than are published, this procedure is conservative. Overall, we have 877 employment series.

For investment, we use published data from the NIPA fixed asset table and the Census ACES survey. We use an annual sample of equipment investment from 1960-2016. For every series we compute the standard deviation of investment over the capital stock, $\frac{I_t}{K_t}$ (information for each series comes from NIPA). We remove the trend from both series using a cubic polynomial. Estimates for the sampling error for each sector come from the Census ACES survey.[64] We make two adjustments

[62] We use cubic polynomials to detrend both series

[63] https://www.bls.gov/web/empsit/cesvarae.htm

[64] https://www.census.gov/data/tables/2015/econ/aces/2015-aces-summary.html. We use information from Table 4c.

Table 12: SLOW CONVERGENCE

Estimating the Relative Bias with $-\hat{\sigma}^2_{\text{SE}}/\hat{\sigma}^2_{\Delta y}$

Employment and Investment

Aggregate	Frequency	Level of Aggregation			
		NAICS 4+	NAICS 3-4	NAICS 2	Aggregate
Employment	Quarterly	-0.534	-0.319	-0.173	-0.0146
Investment	Annual	—	-0.346	-0.170	-0.0176

to these published sampling errors. First, ACES only reports relative sampling errors for the level of investment, while we need sampling errors for $\frac{I_t}{K_t}$. A simple conversion is to multiply the published sampling errors by the mean(I_t/K_t) for each series. This is tantamount to assuming that there is no sampling error in K_t and therefore conservative. Second, we adjust our estimates for depreciation. A simple adjustment is to multiply our sampling errors by $\sqrt{1+\mu^2/[2\delta(2-\delta)]}$ where μ denotes the mean of I_t/K_t and δ is the depreciation rate. We use $\delta = 0.10$. This adjustment makes only a slight difference in practice because typically this constant is quite close to one. In the end, we have 51 investment series.

Table 12 reports how this bias estimate derived in Proposition 4, $-\hat{\sigma}^2_{\text{SE}}/\hat{\sigma}^2_{\Delta y}$ varies with the level of aggregation. Each row reports average bias estimates within the corresponding category. This bias is larger than 50% for employment at the 4+ digit level. No data to obtain an estimate at this level of aggregation is available for investment data. At the 3-4 digit level, the relative bias is above 30% for both aggregates: 32% for employment and 35% for the investment-to-capital ratio. This bias also is relevant (approximately 17%) for both variables at the NAICS 2 or super-sector level (e.g., construction and durables). As was the case for prices, this bias is minimal at the aggregate level.

Summing up, the above results suggest that researchers should be mindful of the missing persistence bias when using sectoral employment and investment data.

F.5 Strategic Complements

In this section we extend the results from Section 3 to the case where prices are strategic complements. Under the Technical Assumptions from Section 2, agents' decision variables are neither strategic complements nor strategic substitutes: when agents adjust they adjust fully to all passed shocks since they have last adjusted. This may not be a reasonable assumption as many authors have argued that strategic complementarities are a central element to match the persistence suggested by VAR evidence (Woodford, 2003; Christiano, Eichebaum and Evans, 1999, 2005; Clarida, Gali and Gertler, 2000; Gopinath and Itskhoki, 2010).

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

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

with $\phi > \rho$ when prices are strategic complements. In line with the strategic complementarity parameters advocated by Woodford, we assume $\phi = 0.944$.

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

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

with $c = (1-\phi)/(1-\rho)$.[65]

Table 13: SLOW CONVERGENCE AND STRATEGIC COMPLEMENTARITIES

$\hat{\rho}$ with Strategic Complementarities

Measure	Calibration	Effective number of agents (N)					
		100	400	1,000	4,000	15,000	40,000
$\hat{\rho}$	Baseline RW: No SC	0.022	0.083	0.182	0.447	0.690	0.787
	Woodford: $\phi = 0.944$	0.005	0.040	0.081	0.258	0.558	0.743
	High SC: $\phi = 0.968$	0.005	0.012	0.044	0.159	0.427	0.631
	Low SC: $\phi = 0.924$	0.011	0.050	0.114	0.325	0.618	0.775
Relative	Baseline RW: No SC	-0.975	-0.904	-0.788	-0.481	-0.198	-0.085
Error:	Woodford: $\phi = 0.944$	-0.995	-0.958	-0.914	-0.727	-0.410	-0.213
$(\hat{\rho}-\rho)/\rho$	High SC: $\phi = 0.968$	-0.995	-0.987	-0.955	-0.835	-0.558	-0.348
	Low SC: $\phi = 0.924$	-0.989	-0.946	-0.876	-0.648	-0.329	-0.159
Median Relative Error:		-0.990	-0.952	-0.895	-0.687	-0.369	-0.186

This table presents results for how adding strategic complementarities to a Calvo model affects the missing persistence bias. Parameter values use in our baseline random walk (RW) calibration: $1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.0616$

Table 13 presents the AR(1) persistence measure, $\hat{\rho}$, in this setting where in all cases we use our random walk calibration (RW) as our baseline: ($1-\rho = 0.14$, $\mu_A = 0.002$, $\sigma_A = 0.0037$, $\rho_I = 1.0$, $\sigma_I = 0.0616$). The first row reproduces the values for the case with no strategic complementarities (this is the same as the third row in our Calvo appendix section: Table 9). The second row presents the case with our baseline level ("Woodford") level of strategic complementarities. This bias is larger with strategic complementarities: With 15,000 units, the relative error is 41% compared to 20% in the case with no strategic complementarities. The main reason for the larger relative error is that $\hat{\rho}_\infty = \phi$. Thus $\phi = 0.86$ without strategic complementarities while its value is higher ($\phi = 0.944$) with strategic complementarities. The other reason convergence is slower its that when strategic complementarities are present and agents adjust, they no longer adjust fully to the aggregate shocks that accumulated since the last time they adjusted. This decreases the strength of the mechanism that recovers the speed of adjustment, namely the covariance of adjustments across agents (see

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

Section 3).[66]

The bottom two rows of Table 13 present robustness for a case with higher (third row) and lower (fourth row) levels of strategic complementarities. Consistent with the intuition provided above, the relative error is higher in the former case for all values of effective N.

G ADDITIONAL INFORMATION

G.1 Estimating IRFs

The section describes in detail how the IRFs displayed in Sections 3.3.1 and 4.1 were constructed. The first method is analytical. As derived in (87), given our assumptions the response of inflation in period $t+k$ to a nominal shock ϵ_t is:

$$E_t\left[\frac{\delta\pi_{t+k}}{\partial\epsilon_t}\right] = (1-\rho)\rho^k.$$

This is shown in the dotted line. The second procedure, shown in the dashed line, uses a simple Monte Carlo ("Simulation") method where the IRF is the response of π to a one grid point increment of Δ of the nominal shock at time t relative to a world where this shock did not occur. In particular, we compute the IRF as

$$E_t\left[\frac{\partial\pi_{t+k}}{\partial\epsilon_t}\right] = (E_t[\pi_t|\epsilon_t = \Delta] - E_t[\pi_t|\epsilon_t = 0])/\Delta.$$

Given that the Monte Carlo method is not polluted by lumpy adjustment if we use the true number of agents in the simulations, the estimated IRFs will not be biased. Finally, we estimate IRFs using both the VAR and MA approaches. They are the solid and dashed-dot lines respectively in Figures 2 and 3, which correspond to the Calvo and Ss models respectively.

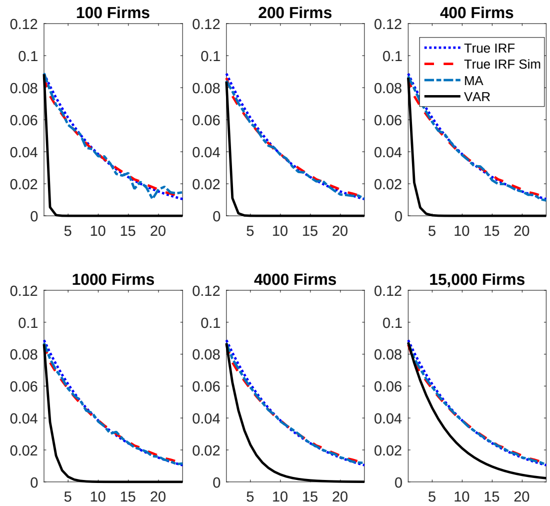
As expected, the Monte Carlo method closely approximates the true response for all N. Two other results jump out. First, this bias is substantial for the VAR approach, particularly for small N. The estimated IRF using this approach is always below the true response. Thus researchers using this approach will infer much faster adjustment to nominal shocks than exists in the model. Second, the MA approach does a good job of estimating the true IRF even in small samples. This suggests that this methodology is a robust way of dealing with the missing persistence bias. Overall, this exercise provides support for using the local projects methodology (Jorda 2005), as it is robust to both misspecification and the missing persistence bias.

G.2 SMM

This section gives details from our SMM Monte Carlo from Section 3.3.2. As a brief, refresher, simulation based estimators are a common way of estimating macroeconomic models because inference only requires the ability to simulate data from the economic model rather than needing to deal with an often analytically intractable or difficult to evaluate likelihood function. Indirect inference is an approach used frequently in this context (Smith, 2008). The goal of indirect inference is to choose the parameters of the economic model so that the simulated model matches closely the observed

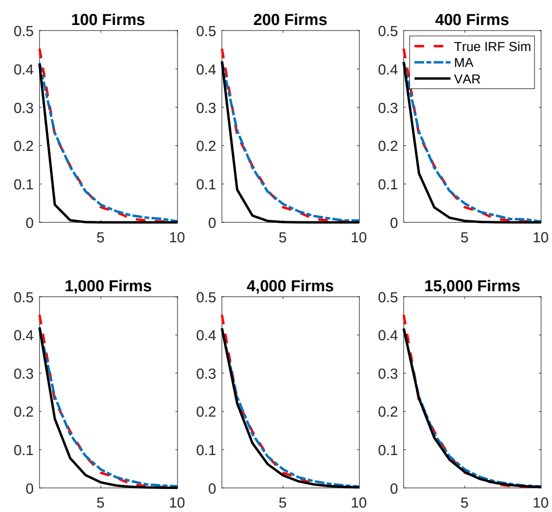
[66] There's a countervailing effect because the firm's own-price-change correlation now is positive. Yet the impact of this effect on aggregate inflation decreases fast as the number of firms grows.

Figure 2: RESPONSE OF INFLATION TO A NOMINAL SHOCK IN A GE CALVO MODEL



This figure shows the IRF of inflation to a nominal shock computed in four separate ways. 1) Using the analytical expression in equation 87 (blue dots); 2) The average (across 100 simulations) of the true non-linear IRF in the model computed via simulation (red dash); 3) Using our MA methodology (light blue dot-dashed) 4) Using our VAR methodology (black solid line). We use the calibration of Nakamura and Steinsson (2010). The parameter values are: $\mu_A = 0.0021$, $\sigma_A = 0.0032$, $\sigma_I = 0.0425$, $\rho_I = 0.66$ and $K = 0.0245$ which implies that $\rho = 0.91$.

Figure 3: RESPONSE OF INFLATION TO A NOMINAL SHOCK IN A GE MENU COST MODEL



This figure shows the IRF of inflation to a nominal shock computed in three separate ways. 1) The average (across 100 simulations) of the true non-linear IRF in the model computed via simulation (red dash); 2) Using our MA methodology (light blue dot-dashed) 3) Using our VAR methodology (black solid line). We use the calibration of Nakamura and Steinsson (2010). The parameter values are: $\mu_A = 0.0021$, $\sigma_A = 0.0032$, $\sigma_I = 0.0425$, $\rho_I = 0.66$ and $K = 0.0245$ which implies that $\rho = 0.91$.

data from the vantage point of some moments or "auxiliary model", which are both informative about the underlying structural parameters and can easily be computed in both the model and the data. The parameters of the underlying economic model are then chosen so as to minimize the difference between the parameter estimates of the auxiliary model in the model and in the data. Under mild assumptions, this approach will identify the structural parameters of interest.

A common form of indirect inference is IRF matching. In the language of indirect inference, the auxiliary model is the IRFs and one uses this auxiliary model to estimate the structural parameters of a model. Section 3.3.2 show that while indirect inference has many virtues, this methodology must be applied with care if the missing persistence bias is present. When an underlying variable has lumpy adjustment and IRFs are estimated using the "VAR" approach, the estimates of the IRF will be biased. This bias in the estimation of the auxiliary equation can translate into bias in the estimates of the underlying structural parameters.

One solution to this issue is to estimate IRFs using a methodology that is robust to the missing persistence bias such as Jorda (2005). A more general solution is to simulate data in exactly the same form as the researcher has access to in reality. In particular, it is crucial to use actual sample sizes when estimating the auxiliary model: if the researcher simulates much larger samples of data in the model then one would eliminate the missing persistence bias in the model but not in the data, potentially biasing the estimates of the parameters of interest.

Table 14: SMM TABLE

Monte Carlo example: matching IRFs by simulated method of moments (SMM)

	Estimator	Weight Matrix	**Model moments** Effective number of agents (N) in simulations: 400	1,000	4,000	15,000
Data	VAR	Identity	0.250	0.730	0.820	0.840
($N = 400$)		Proportional	0.250	0.510	0.760	0.770
($1-\rho = 0.25$)		Optimal	0.250	0.710	0.820	0.840
Data	MA	Identity	0.250	0.250	0.250	0.250
($N = 400$)		Proportional	0.250	0.250	0.250	0.250
($1-\rho = 0.25$)		Optimal	0.250	0.250	0.250	0.250

This table documents that it is important to treat real and simulated data similarly when the missing persistence bias is present using a simple Monte-Carlo. The number of underlying agents is 400 in the "Data". We compute the IRF of inflation to a nominal shock in two ways: the VAR approach (top panel) and MA approach (bottom panel). The true frequency of adjustment, $1-\rho = 0.25$. We compute the analogous model implied IRF by simulation. The only difference across the simulations is the number of underlying agents used to calculate this IRF. All rows show the estimated $1-\hat{\rho}$ from the SMM estimation and all results are averages across 100 simulations.

Table 14 illustrates this point for a simple Monte Carlo simulation that builds on our previous Calvo model. Consider an applied researcher who wants to estimate the frequency of adjustment (the structural parameter) by SMM using the impulse response function of inflation to a nominal shock as the auxiliary model. This IRF is a sensible choice since the k^{th} element of the IRF is equal to $\rho^k(1-\rho)$. This is a highly stylized example – in more complicated frameworks this IRF would depend on more than one structural parameter. The example is kept deliberately simple to illustrate the main point.

We assume that there are 400 price setting firms in the data who all use Calvo pricing with the same frequency of adjustment, $1-\rho$, equal to 0.25. The data moment is the IRF of inflation to a nominal shock computed in this model. Motivated by Proposition A.10, which shows that the element upon impact of the IRF, ($k=0$), is not biased under any estimation methodology (this is clear from Figure 1 in the main text), we match the 1st to 12th lag elements of the IRF.

The top panel of Table 14 illustrates the case when both the data and model IRFs are computed using the standard VAR approach. Each row shows the results from the SMM estimation for three different weight matrices: "optimal" (inverse of the variance-covariance matrix of the data moments), "proportional" (inversely proportional to moment size) and "identify" (equal weights). Each column varies the number of underlying firms when the researcher estimates the IRF. In all cases we compute averages of the model moments across 100 simulations. Two results are clear. The first column shows that the SMM estimator provides an unbiased estimator of the frequency of adjustment when the researcher's simulation has the same number of firms in the model as are in the data. This gives support for the folk wisdom that researchers should treat real and simulated data similarly.

The perils of not doing this are shown in the other three columns. Since the underlying data has 400 firms, the missing persistence bias is severe. If a researcher tried to match this IRF using a simulation with 15,000 firms, she would infer a much faster speed of adjustment as shown by the last column of Table 14. The reason is that the VAR approach is subject to the missing persistence bias and this bias diminishes with the number of effective firms (compare the top left panel of Figure 1, which shows the IRF for 100 firms to the bottom right panel which shows the IRF with 15,000 firms). The only way to match the biased data estimate with an unbiased estimate is by increasing the frequency of adjustment – this is why the estimated frequency increases as one moves from left to right across the columns. In contrast, the bottom panel shows that no such issue exists if IRFs are estimated by the MA approach. This is because this approach is immune to the missing persistence bias.

G.3 A Simple Method for Approximating the Entire IRF in General *Ss* models

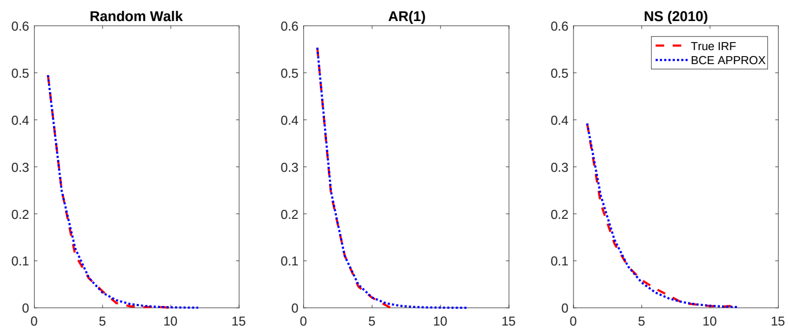
As discussed in Appendix A.2, for small σ_A and an infinite number of agents, we may approximate approximate aggregate inflation in a general Ss model by a distributed lag of aggregate shocks:

$$\Delta y_t^{\infty} \simeq \sum_{k\geq 0} I_k v_{t-k}^A.$$

As argued in that appendix, $(I_k)_{k\geq 0}$ will be a good approximation for the corresponding IRF for an aggregate with any number of effective agents, N.

In contrast with the Calvo model, for an Ss model I_0 no longer is equal to the average fraction of adjusters but larger (Caballero and Engel, 2007). The reason is that with Ss models, the response of aggregate inflation to a positive shock upon impact will be the sum of two components. The first component ('intensive margin') is the contribution to aggregate inflation of agents that would have adjusted with or without the impulse. Agents that were planning to increase their prices do so by a bit more and agents that were planning to reduce their prices do so by a bit less. The second component ('extensive margin') captures agents that change their decision on whether to adjust their price in response to the impulse. Some were planning to remain inactive but end up increasing their price, others were planning to decrease their price but end up remaining inactive. The first component is equal to the fraction of adjusters, $1-\rho$, and is the same in Ss and Calvo models that match this moment. The second component is not present in the Calvo model while it is strictly positive for Ss model where agents' adjustments are triggered both by idiosyncratic and aggregate shocks. In this dimension, the particular Ss model we studied in Appendix A2 is closer to the Calvo model, since the second component described above is not present as shocks are triggered only by idiosyncratic shocks.

Figure 4: TRUE VERSUS APPROXIMATE IRFS IN Ss MODELS



This figure shows true and approximate IRF of inflation to a nominal shock in three calibrations of our general Ss model: 1) Idiosyncratic shocks follow a Random walk 2) Idiosyncratic shocks follow an AR(1) and 3) Nakamura and Steinsson 2010. In all cases our approximation to the true IRF works well. See Table 11 for calibration details.

The IRFs of all simulated models was found to decrease at an approximately geometric rate. This suggests approximating the true IRF of an Ss model by the IRF of a Calvo model where the fraction of adjusters is I_0 instead of $1-\rho$. This leads to

$$I_k \simeq (1-I_0)^k I_0$$

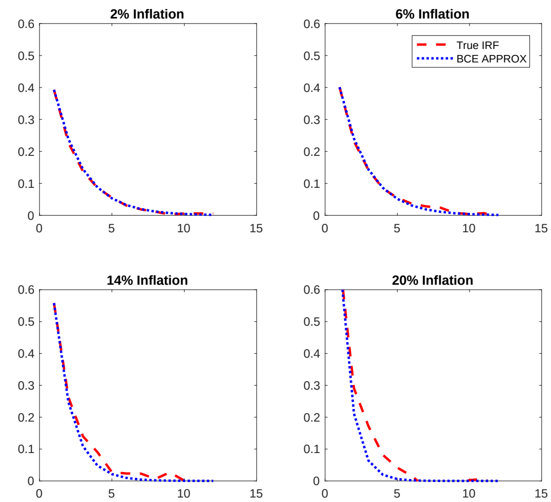
and we can use the expression in Proposition 1 to estimate the magnitude of this bias, with I_0 in the role of $1-\rho$. This approximation helps explain why even though Ss and Calvo models lead to a significant missing persistence bias for inflation for sectoral data, this bias is larger for Calvo models than for Ss models (see Appendix F).

Figure 4 shows actual IRFs and IRFs obtained via this approximation for three Ss models. Shocks follow a random walk in the first model and AR(1) processes In the second and third model. The third model is from Nakamura and Steinsson (2010). See section F.3for the calibration details. The approximation for the IRF works remarkably well in all cases.

It should be noted, though, that the above approximation breaks down when the mean of aggregate shocks is large as illustrated in Figure 5, that plots the actual IRF and the above approximation for the Nakamura and Steinsson (2010) model in three scenarios that only differ in the values of core inflation. This figure suggests that the approximation stops being good when annual core inflation is above 10%.[67]

[67] This conclusion is likely to be conservative, since a well established empirical fact is that core inflation and inflation volatility are positively correlated, so by the time the economy's core inflation reaches 10%, the variance of shocks is likely to be larger as well, which makes the approximation less imprecise.

Figure 5: QUALITY OF IRF APPROXIMATION VARIES WITH TREND INFLATION



This figure shows how the quality of our approximation of IRF of inflation to a nominal shock varies with different levels of trend inflation (more generally, depreciation). We use the calibration of NS (2010) and all that varies across the labels is the level of annual trend inflation. Looking across the four panels, it is clear that the approximation works well for low levels of trend inflation but breaks down for high levels.

G.4 Implications for Estimating the New Keynesian Phillips Curve

In this section we discuss the implications of the missing persistence bias for estimation of the New Keynesian Phillips curve (NKPC). As will become clear, it depends on what method one uses to do the estimation (e.g. GMM), which NKPC one estimates (purely forward looking or whether backward looking terms are included) and what is in the information set. However, for most of the common specifications where backward looking terms are included or lagged inflation rates are used as instruments, the missing persistence bias may be relevant.

To see this consider the basic forward looking NKPC with a backward looking term:

$$\pi_t = \lambda \mathrm{mc}_t + \beta E_t[\pi_{t+1}] + \gamma \pi_{t-1} \tag{110}$$

There are two prominent methods of estimating equation (1): GMM and NLS

GMM

In an important paper. Gali and Gertler (1999) provide a methodology for estimating the NKPC. They do so by noting that equation (1) can be re-written as:

$$\pi_t = \lambda \mathrm{mc}_t + \beta \pi_{t+1} + \gamma \pi_{t-1} + \varepsilon_{t+1} \tag{111}$$

where π_{t+1} is realized inflation in period $t+1$ and ε_{t+1} is an expectational error. Under rational expectations, the error in the forecast of inflation in $t+1$ is uncorrelated with information dated t and earlier, thus β in equation (2) can be consistently estimated using information from variables dated t and earlier. A typical example of this approach is Gali et al. (2005). They use GMM with four lags of inflation, two lags of the labor income share, the output gap and wage inflation as instruments.[68] Because the NKPC is linear, GMM is equivalent to 2SLS with a certain weighting matrix.[69] Thus we know that the point estimates will be identical to the following two-step procedure (though the standard errors of this two-step procedure are less efficient than GMM):

1. Regress π_{t+1} on the instrument set Z_t then keep the predicted inflation rate implied by this regression: $\hat{\pi}_{t+1}$.
2. Substitute $\hat{\pi}_{t+1}$ into our NKPC estimating equation (equation 2) and estimate it by OLS. The coefficient on $\hat{\pi}_{t+1}$ is the main coefficient of interest.

The missing persistence bias may affect these estimates through two channels if lagged inflation is used in the instrument set. The first problem is that the first stage will be biased downward thus affecting the predicted regressor, $\hat{\pi}_{t+1}$, leading to downward biased estimates of $\hat{\beta}$. Second, if backward inflation terms are included in the NKPC (as above) then all the coefficients in the second stage regression will be biased.

NLS

Linde (2005) proposes a different way to estimate the NKPC. He starts imposing rational expectations. This means that we can write inflation as: $\pi_t = E_t[\pi_{t+1}] + \varepsilon_{t+1}$ where ε_{t+1} is orthogonal to the information set in period t. Plugging this into the original NKPC and rearranging gives:

$$\pi_{t+1} = \frac{1}{\beta}\pi_t - \frac{\lambda}{\beta}\mathrm{mc}_t - \frac{\gamma}{\beta}\pi_{t-1} + \varepsilon_{t+1} \tag{112}$$

which can be estimated by OLS or NLS to recover the parameters. Clearly this approach to estimating the NKPC would be affected by the missing persistence bias if lagged inflation terms (e.g. backward looking terms) are included in the specification, which they commonly are.

Table 15 uses simulated data from our GE Calvo model to illustrate how the estimates of the NKPC are affected by the missing persistence bias. The model that is estimated is the standard forward looking NKPC:

$$\pi_t = \lambda \mathrm{mc}_t + \beta\pi_{t+1} + \varepsilon_{t+1}.$$

We consider three methods for estimating the NKPC: GMM, 2SLS and the two-step procedure discussed above. We use four lags of inflation and output as instruments. All simulations use $T = 1000$ and all that varies across simulations is N. The results shown are medians across 100 simulations.

Irrespective of the estimation method used, the missing persistence bias affects the estimates of the NKPC. This can be seen by noting that the coefficients for both $\hat{\beta}$ and $\hat{\gamma}$ are uniformly increasing in N. This bias can be severe when N is small. For example, the estimated $\hat{\beta}$ when $N = 100$ is 33-50% lower than the estimated $\hat{\beta}$ when $N = 15,000$. This downward bias is in the 10-25% range when $N = 400$. This suggests that caution should be used when estimating NKPCs with sectoral inflation data if you are using lagged inflation as an instrument.

[68] The instrument set in Gali and Gertler (1999) is similar. It consists of four lags each of price inflation, the labor share of income, the output gap, the spread between long and short interest rates, compensation growth, and commodity price inflation.

[69] Inverse variance-covariance matrix of the instruments.

Table 15: NEW KEYNESIAN PHILLIPS CURVE ESTIMATION

Estimation	Parameter	Effective number of agents (N)				
		100	400	1,000	4,000	15,000
GMM	$\hat{\beta}$	0.531	0.804	0.969	1.053	1.071
	$\hat{\gamma}$	-0.476	-0.227	-0.104	-0.029	-0.019
2SLS	$\hat{\beta}$	0.690	0.956	1.032	1.078	1.088
	$\hat{\gamma}$	-0.323	-0.109	-0.043	-0.008	-0.001
Two-step Procedure	$\hat{\beta}$	0.794	1.071	1.117	1.178	1.191
	$\hat{\gamma}$	-0.297	-0.061	0.016	0.068	0.085

This table documents how the number of underlying observations can effect estimation of the NKPC. Throughout we use the calibration of NS (2010): ($1-\rho = 0.09$, $\mu_A = 0.0021$, $\sigma_A = 0.0037$, $\rho_I = 0.66$, $\sigma_I = 0.0515$). The model that is estimated is the standard forward looking NKPC: $\pi_t = \lambda \text{mc}_t + \beta\pi_{t+1} + \varepsilon_{t+1}$. We consider three methods for estimating the NKPC: GMM, 2SLS and the two-step procedure discussed in the text. We use four lags of inflation and output as instruments. All simulations use $T = 1000$; all the varies across simulations is N. Results shown are medians across 100 simulations.

To sum up, both the GMM and NLS methods are potentially affected by the missing persistence bias if backward looking terms are included and the GMM approach is affected in past inflation is used in the instrument set.

G.5 Reset Price Inflation

The basic idea behind reset price inflation is to make inferences about the underlying shocks using information contained only in observed price changes where the implicit assumption is that when a firm adjusts it is adjusting (“resetting") to its optimal price. Specifically, define $p_{i,t}$ as the log price of item i and time t and define a price change indicator as:

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

The reset price, $p_{i,t}^{\text{reset}}$, for prices that do not change is simply the current price. The reset price for non-changers is then updated using the rate of reset price inflation estimated from the price changers in the current period:

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

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

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

where $\omega_{i,t}$ denote i's relative expenditure weight at time t. Thus reset price inflation is the "inflation rate" conditional on the price adjustment. With Calvo price setting and assuming that the technical assumptions from Section 2 hold, it is easy to show that reset price inflation reduces to the following formula:[70]

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

This justifies using reset price inflation as an estimate of sectoral shocks. Next we present simulation results showing that reset price inflation is also a good method to recover the true shock innovations in both more realistic Calvo environments with large idiosyncratic shocks and Ss-type settings.[71]

Table 16: DOES RESET PRICE INFLATION RECOVER THE TRUE SHOCKS?

REGRESSION OF ESTIMATED SHOCK ON TRUE SHOCK: RESET PRICE INFLATION

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

Monte-Carlo evidence: do we recover the true shock in practice?

In order to verify that our shock measure recovers the true shock, we simulate both a Calvo and an *Ss* model with the following standard parameter values: the frequency of adjustment = 0.2, μ_{agg} = 0.002, σ_{agg} = 0.003, ρ_I = 0.97, σ_I = 0.04 (we also tried something farther from a random walk: ρ_I =0.7). These economies were simulated for T=300 periods with a burn in of 100 periods. Notice that there are two types of shocks: aggregate shocks that affect everyone and idiosyncratic shocks that are firm specific. In each simulation we ran the following regression:

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

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

[71] We also estimated the shocks using a repeat-price-change approach (similar to the Case-Shiller index) and found similar results.

where v_t is our shock measure (reset price inflation) and z_t is the true shock innovation from each simulation. The level and fit of this regression is informative of how well our shock measure proxies for the true shock. It is an important robustness check because we want to make sure that we can recover an unbiased estimate of the true aggregate shock in a situation where idiosyncratic shocks are realistically large relative to aggregate shocks. The results (averaged across 100 simulations) are comforting and shown in Table G.5.

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

G.6 Relevance of the Missing Persistence Bias for Prices

In section 5 we assessed the relevance of the missing persistence bias for the US CPI using our baseline monthly sample. In this appendix we assess the relevance of this bias for the entire bi-monthly sample. The results are shown in Table 17. We include the bootstrapped results from our baseline sample for comparison.

Two results are clear. First, the magnitude of this bias is similar across both samples. Second, the missing persistence bias is substantial for $N < 1000$ reenforcing our conclusion that researchers should be careful when using sectoral data to estimate persistence.

The bi-monthly sample also allow us to compute $\hat{\rho}^N$ for a larger N ($N = 40,000$) than is possible with just the monthly sample. Consistent with our prediction that the missing persistence bias decreases with N, we find that $\hat{\rho}^{40,000}$, which is equal to 0.345, is larger than $\hat{\rho}^{15,000}$ computed using either sample, which is equal to 0.316 or 0.328, respectively.

Table 17: ESTIMATING THE MISSING PERSISTENCE BIAS: INFLATION

Measure	Source		Effective number of agents			
		100	400	1,000	4,000	15,000
Relative bias:	CPI monthly (bootstrap)	−0.845	−0.615	−0.394	−0.124	−0.042
	CPI bi-monthly (bootstrap)	*-0.864*	*-0.635*	*-0.420*	*-0.162*	*-0.049*
Estimate for $\hat{\rho}^N$:	CPI monthly (bootstrap)	0.051	0.127	0.200	0.289	0.316
	CPI bi-monthly (bootstrap)	*0.047*	*0.126*	*0.200*	*0.289*	*0.328*

G.7 Application #1: A Simple Test of the Calvo Model

This section provides details for our first application from Section 6.1. As noted in the main text, we started with this example because (i) the assumptions in the BK paper are identical to those underlying the results in Section 3 (ii) it highlights that the missing persistence bias is relevant in U.S. pricing data at the sectoral level and (iii) we are able to calculate the exact magnitude of this bias in this case from the CPI micro database.

BK conduct a simple test of the Calvo model using CPI microdata. They start by using the micro data to estimate the frequency of price adjustment in each sector, λ_s. Next, they estimate the following regression by OLS:

$$\pi_{st} = \rho_s \pi_{s,t-1} + e_{st}, \tag{113}$$

where π_{st} is inflation in sector s at time t. Under the assumptions of the Calvo pricing model considered in Section 3 with $N = \infty$, we should find that $\hat{\rho}_s$ is approximately equal to $1-\hat{\lambda}_s$. In contrast, BK find that in all sectors $\hat{\rho}_s$ is substantially smaller than $1-\hat{\lambda}_s$ and interpret this as strong evidence against the Calvo model.

We test whether the missing persistence bias is responsible for BK's result using the bias correction approach outlined in section 3.2. We construct our proxy for v_{st} using the reset price inflation methodology of Bils, Klenow and Malin (2012).

In constructing estimates of π_{st} and v_{st}, 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 this 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.[72] This leaves us with 66 sectors.

Once we have our 66 reset price inflation estimates, we implement our bias correction procedure by including our measure of the sectoral shock, v_{st}, as an additional control in equation (113)

$$\pi_{st} = \beta_s \pi_{s,t-1} + \gamma_s v_{st} + e_{st}. \tag{114}$$

Proposition 2 from Section 3.2 implies that if we estimate β_s and γ_s in the above equation without imposing any constraints across them, then $\widehat{\gamma}_s$ will be an unbiased estimate of the actual fraction of adjustment λ_s. We then examine how close $\widehat{\gamma}_s$ is to λ_s.

As a first step we replicate BK's results using our 66 sectors. In particular, we estimate equation (113) using the micro data, and denote the implied frequency of adjustment estimates as $\lambda_s^{\text{VAR}} = 1-\widehat{\beta}_s$. As in BK, we find that $\widehat{\beta}_s \ll 1-\lambda_s^{\text{micro}}$, where λ_s^{micro} denotes the true frequency of adjustment, estimated from the micro level quote-lines. Across all 66 sectors, the mean (median) estimate of $\widehat{\beta}_s$ is 0.08 (0.06) compared to 0.88 (0.93) for $1-\lambda_s^{\text{micro}}$ and $\widehat{\beta}_s < 1-\lambda_s^{\text{micro}}$ in all sectors, with the exception of only one. Now that we have established that BK's baseline result holds in our dataset, we implement our bias correction procedure by estimating equation (114) using our constructed shock measure, v_{st}.

We start with some definitions. Denote the coefficient on our sectoral reset price inflation measure by $\lambda_s^c = \widehat{\gamma}_s$, where the superindex c stands for "corrected". Define $\lambda_s^{\text{VAR}} = 1-\widehat{\beta}_s$ where $\widehat{\beta}_s$ is estimated using equation (113). To gauge the extent to which the λ_s^c correct the missing persistence bias, we regress the change in estimated speed of adjustment we achieve in a given sector, $\lambda_s^c - \lambda_s^{\text{VAR}}$, on the magnitude of this bias, $\lambda_s^{\text{micro}} - \lambda_s^{\text{VAR}}$. That is, since we are in a rare situation where we actually know this bias, we are able to estimate by OLS the following equation:

$$(\lambda_s^c - \lambda_s^{\text{VAR}}) = \alpha + \eta \text{bias}_s + \epsilon_s, \tag{115}$$

with $\text{bias}_s \equiv \lambda_s^{\text{micro}} - \lambda_s^{\text{VAR}}$. Here η is the coefficient of interest as it captures the extent to which our bias correction actually decreases this bias. If this bias reduction is large but unrelated to the magnitude of this bias, the estimated value of α will be large while η won't be significantly different from zero. By contrast, if this bias reduction is proportional to the actual bias, we expect an estimate of η that is significantly positive, taking values close to one if this bias completely disappears.

The results are implemented in the main text in Table 4. The results for our simulated Calvo and *Ss* models are calibrated multi-sector versions of the model discussed in Appendix E.1 These

[72] We only chose those sectors for which we could have data for the entire sample period because we want to have a large T. Overall, we use data from 1988m2-2007m12, or $T = 238$.

Table 18: RECREATION OF BK (2004) TABLE 4

	Mean ρ_i	Correlation between ρ_i and λ_i	Obs
BK 2004	-0.05	0.26	123
BCE 2018	0.08	0.25	66
BCE 2018 ($N_i < N_{mean}$ and $\lambda_i < \lambda_{mean}$)	0.04	0.51	32
BCE 2018 ($N_i < N_{mean}$ and $\lambda_i > \lambda_{mean}$)	0.07	0.46	9
BCE 2018 ($N_i > N_{mean}$ and $\lambda_i < \lambda_{mean}$)	0.16	0.17	15
BCE 2018 ($N_i > N_{mean}$ and $\lambda_i > \lambda_{mean}$)	0.17	0.01	10

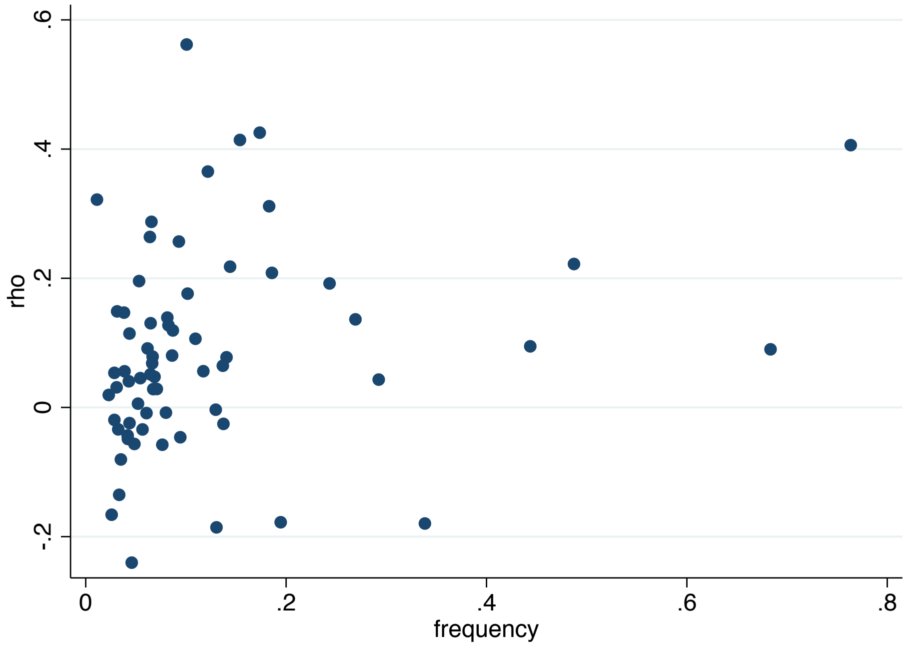
multi-sector models provide a useful laboratory to test in a controlled setting whether the missing persistence bias is relevant and whether our bias correction approach works. The full calibration results are given in Appendix E.2. Since a crucial element in these calibration is to work with the correct number of price setters in each sector, we set the number of effective price-setters in each sector equal to the number of effective price-setters in the relevant sector of the CPI microdata. In particular, we use item level expenditure weights w_i, $i = 1, 2, .., n$, with $w_i > 0$ and $\sum_{i=1}^{n} w_i = 1$ within each sector. Then the effective number of units in each sector, N_s, is definied as the inverse of the Herfindahl index:

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

Next we reproduce the lower part of Table 4 in BK on the correlation between the estimated ρ_i and the micro based λ_i, the frequency of adjustment, where i denotes sector. Both Table 4 and Figure 2 in BK are remarkable: there appears no correlation whatsoever. The first row reproduces the results from BK (2004). It shows that the mean ρ_i is very close to zero and that the correlation between the estimated ρ_i and the micro based λ_i is somewhat positive (0.26) despite the fact that the Calvo model predicts that these two objects should be perfectly negatively correlated. The second row shows results using our sample. Consistent with BK, we find that the mean ρ_i is also very close to zero 0.08. Furthermore, the correlation between ρ_i and λ_i is also slightly positive (0.25). Thus, we find almost identical results to BK despite our slightly different data sample.

The basic Calvo model predicts that there should be a negative relationship even with small N (we find this to be true in our simulations) so there must be some pattern in the number of observations across sectors, N_i, that explains this pattern. In our sample, ρ_i and λ_i are both positively correlated with N_i. This means that some sectors are really affected by the missing persistence bias because they have both low N_i and low λ_i. The bottom four rows of the table highlight this by comparing results for four cases: i) those with N_i and λ_i below their means ii) those with N_i below its mean and λ_i above iii) those with N_i above the mean and λ_i below iv) those with both N_i and λ_i above the mean. While the sample sizes are small, it is striking to compare case (i) to case (iv). It is clear that when a sector has a lot of observations and a high frequency of adjustment, we get much closer to the negative correlation that one would expect in theory. In fact, it turns out that the sector with the highest frequency is a bit of an outlier in that it has both a high estimated regression coefficient ρ, a large N_i and a high frequency of adjustment. If you exclude that one sector, then for case (iv) the correlation falls from 0.01 to -0.34. This is consistent with the results in table 5 in the main text, which shows that this bias ($\lambda_s^{\text{VAR}} - \lambda_s^{\text{micro}}$) is decreasing in the frequency of adjustment and the number of observations.

Figure 6: RECREATION OF BK TABLE 4 IN FIGURE FORM



G.8 Application #2: Does Inflation Respond More Quickly to Sectoral Shocks?

This section describes the details our of second application. To understand BGM's approach, we must first introduce some terminology. Define Π_t as a column vector with monthly sectoral inflation rates in period t, for sectors 1 through S, where S denotes the number of sectors. BGM assume that Π_t can be decomposed into the sum of a small number R of common factors, C_t, and a sectoral component, e_t:

$$\Pi_t = \Lambda C_t + e_t, \tag{116}$$

where Λ denotes an SxR matrix of factor loadings that are allowed to differ across sectors, while C_t and e_t are Rx1 and Sx1 matrices. This formulation allows them to disentangle the fluctuations in sectoral inflation rates due to the macroeconomic factors—represented by the common components C_t with sector specific weights—from those due to sector-specific conditions represented by the term e_t.

BGM extract R principal components from the large data set Π_t to obtain consistent estimates of the common factors.[73] Next, they regress each sectoral inflation series on these common factors,[74] denoting the predicted aggregate component, $\lambda_i' C_t$, by π_{st}^{agg}, and the residual that captures the sector-specific component, e_{st}, by π_{st}^{sect}. This methodology decomposes each sectoral inflation series into aggregate and sectoral components that are orthogonal:

$$\pi_{st} = \lambda_s' C_t + e_{st} = \pi_{st}^{\text{agg}} + \pi_{st}^{\text{sect}}. \tag{117}$$

[73] 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.

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

To calculate IRFs with respect to the common and sectoral shocks, BGM fit separate AR(13) processes to the π_{st}^{agg} and π_{st}^{sect} series and measure the persistence of shocks by the sum of the 13 AR coefficients. This is a standard method for estimating IRFs and is motivated by the observation that if there is a lot of persistence in the data then the sum of the AR coefficients should be close to one. For example, if the underlying microdata were generated by a Calvo model with $N = \infty$, then this sum is equal to one minus the frequency of adjustment. Decreases in the adjustment frequency increase actual persistence and this method of measuring IRFs reflects this accurately.

We start by reproducing BGM's benchmark results using the CPI data. There are a few differences between our sample and BGM's.[75] The first two columns show results for BGM's baseline sample taken directly from Table 1 in their paper. The third and fourth columns show results using BGM's methodology on the data sample that is closest to our setting: using only PCE inflation series to construct the aggregate factors (Equation 116) and the 1988-2005 time period. The last two columns show our results when we implemented BGM's methodology with CPI data.

Table 19: BGM (2009): ESTIMATED PERSISTENCE TO AGGREGATE AND SECTORAL SHOCKS

Sum of AR coefficients for AR(13)

	BGM Sample (Baseline)		BGM Sample (PCE + 88-05)		BLS Sample (CPI + 88-07)	
	π_{st}^{agg}	π_{st}^{sect}	π_{st}^{agg}	π_{st}^{sect}	π_{st}^{agg}	π_{st}^{sect}
Mean	0.92	−0.07	0.58	−0.02	0.45	−0.11
Median	0.94	−0.01	0.66	0.09	0.64	−0.04

Table 19 shows that despite differences in the data used, we find similar results to BGM when we replicate their methodology with CPI data.[76] In all cases there is clear evidence of significant persistence to aggregate shocks and negligible persistence to sectoral shocks. While the amount of persistence to aggregate shocks is smaller in the CPI relative to BGM's baseline, a comparison between the third and fifth columns shows that these differences disappear once we use similar underlying data and time periods.[77] Overall, then, BGM's methodology robustly delivers the result that inflation responds faster to sectoral than aggregate shocks. However, given that price adjustment is lumpy and sample sizes are small for the sectoral series, the missing persistence bias could also explain this result. We explore this possibility next.

To implement the MA approach we need estimates of both aggregate, m_t, and sectoral shocks, x_{st}, for each sector s. To get each we use our sectoral reset price shock measures, v_{st}'s described in Appendix G.5. These were computed from CPI microdata over the period 1988:03-2007:12. Define V_t as the Sx1 vector with the period t sectoral shock measures. Our proxy for aggregate shocks is the first R principal components of V, denoted by m_t^k, $k = 1, 2, ..., R$. The logic for this approach is

[75] First, BGM use information on both prices and quantities whereas we just use information on prices. Second, BGM use a longer sample period (1976-2005) than we have (1988-2007). Finally, BGM use more data (BGM use 653 series, half of which are price series) whereas we use 66.

[76] We report results that assume there are 5 common factors.

[77] Reassuringly, Mackoviak, Moench and Wiederholt (2011) reach a similar to conclusion to BGM using the CPI data and a different methodology.

that aggregate shocks are the common component of the υ_{st}'s since by definition they affect each of these series.

We compute the pure sectoral shock as a residual. In particular, we decompose υ_{st} into the sum of an aggregate and a sectoral component and we recover the sectoral shocks by regressing each sectoral reset price series on our estimated aggregate shocks. Since we are using retail data, we include lags of the aggregate shocks in order to allow for some delay in these shocks propagating up the supply chain. Denote the pure sectoral shock as x_{st}.[78] Concretely:

$$\upsilon_{st} = \sum_{k=1}^{R} \sum_{j=0}^{J} \gamma_{sj}^{k} m_{t-j}^{k} + x_{st}, \tag{118}$$

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

Now that we have our R aggregate shocks, m_t^k, and a sectoral shock, x_{st}, for each of our 66 sectors, we can implement our MA approach to estimate IRFs. We do this by regressing each sectoral inflation series on distributed lags of the aggregate and sectoral shocks:

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

where $\eta_s^k(L) = \sum_{j\geq 0} \eta_{sj} L^j$ and $\nu_s(L) = \sum_{j\geq 0} \nu_{sj} L^j$ denote lag polynomials. In order to parsimoniously estimate these lag polynomials, we model each $\eta_s^k(L)$ and $\nu_s(L)$ as quotients of two second degree polynomials.[79] This allows us to flexibly approximate a variety of possible shapes for our IRFs while also maintaining parsimony.

Table 20 shows that our baseline results from Section 6.2 are robust to reasonable variations in the order of these polynomials and in using local projection directly. In particular, the only difference between table 20 (below) and Table 5 in the main text is that the latter used 2 AR and 2 MA lags, while the former used 1 AR lag and 3 MA lags. Regardless on what polynomials one uses, we find similar results: after correcting for the missing persistence bias, we no longer find strong evidence that sectoral inflation responds differently to aggregate and sectoral shocks.

G.8 provides Monte Carlo evidence that our procedure works well in practice.

Monte-Carlo Evidence that the Methodology Works

Next we verify that the methodology we proposed in Section 5.2 for recovering the persistence of sectoral inflation to aggregate and sectoral shocks is an improvement over the standard VAR methodology, which is subject to the missing persistence bias. We test this using a multi-sector Calvo model as a laboratory with both aggregate and sectoral shocks. In this model, the assumptions of Section 3.1 hold so that we know that for a given frequency of adjustment (1-ρ), the estimated response time is equal to $\frac{\rho}{1-\rho}$ to *both* aggregate and sectoral shocks. In other words, in this model we know both what the true level of persistence is and that it is the same to both aggregate and sectoral shocks.

[78] Our results are robust to ignoring these distributed lags of common components yet we believe it is more realistic to include them so they are including in our baseline.

[79] We do not have enough data to estimate an unrestricted version of this equation given that we only have 238 observations for each series and R is the number of lags in each lag polynomial coefficients.

Table 20: THE RESPONSE OF SECTORAL INFLATION RATES TO AGGREGATE AND SECTORAL SHOCKS

Median of estimated expected response times to shocks

	2 PCs		4 PCs		6 PCs	
nlags	agg	sec	agg	sec	agg	sec
0	3.05	3.04	2.42	1.81	2.22	1.50
	(0.91)	(0.54)	(0.34)	(0.67)	(0.33)	(0.63)
3	3.20	2.70	2.06	2.06	2.29	1.53
	(0.86)	(0.59)	(0.45)	(0.70)	(0.38)	(0.64)
6	2.65	3.42	2.70	2.90	2.57	1.70
	(0.68)	(0.70)	(0.52)	(0.56)	(0.28)	(0.58)
12	2.84	1.76	2.76	2.82	2.90	1.44
	(0.48)	(0.54)	(0.48)	(0.48)	(0.25)	(0.59)

In order to be consistent with our previous work we use our baseline Calvo calibration where $\mu_A = 0.003$, $\sigma_A = 0.0054$, and $\sigma_I = 0.048$ and $\rho = 0.86$. We also consider a second calibration with a higher frequency of adjustment ($\rho = 0.80$) in order to show that our results work for a variety of frequencies. We then simulate data from a version of this model that has 50 sectors, with 200 firms and 1000 periods per sector. We then implement the two methodologies discussed in Section 5.2. using this simulated data. In particular, we estimate the persistence of sectoral inflation, π_{st} to both aggregate and sectoral shocks. We use the estimated response time as our measure of persistence since we know it's exact value in our simulations and it is what we reported in Table 21. We run this experiment 100 times and average across simulations.

Table 21: COMPARING METHODS FOR RECOVERING PERSISTENCE

	VAR		BEC		Theory	
	Agg	Sec	Agg	Sec	Agg	Sec
$\rho = 0.86$						
Mean	5.090	1.345	5.779	6.082	6.143	6.143
Median	5.090	1.334	5.843	6.076	6.143	6.143
Std. Deviation	0.000	0.139	0.249	0.033	0.000	0.000
$\rho = 0.80$						
Mean	3.853	1.576	4.026	4.051	4.000	4.000
Median	3.853	1.563	4.029	4.056	4.000	4.000
Std. Deviation	0.000	0.143	0.024	0.033	0.000	0.000

This table documents how different methods of estimating persistence do at recovering the true persistence to nominal shocks. We consider two methodologies: the standard VAR methodology and the one described in Section 5.2 of this paper (BEC). The measure of persistence is the expected response time, which under the assumptions of Section 3.1 (Calvo assumptions) is equal to $\frac{\rho}{1-\rho}$. We consider two calibrations. The first (baseline) uses the same parameter values as our baseline Calvo calibration ($\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$ and $\rho = 0.86$). The second calibration uses the same parameter values except for $\rho = 0.80$.

The results are shown in Table 21. The last two columns ("Theory"), show what the true level of persistence in the model. This is equal to $6.14 = \frac{0.86}{0.14}$ in the first calibration and $4.00 = \frac{0.80}{0.20}$ in the second. The first two columns show the results from using VAR's methodology while the second two columns show the results from using our methodology. Two results stick out. Comparing (BEC) to (VAR), we see that our methodology (BEC) does a good job of recovering the true level of persistence to both aggregate and sectoral shocks. The estimated level of persistence to both shocks are (a) similar to each other and (b) close to the true value. This is not true if one uses the VAR methodology. In this case one would infer that inflation responds much more slowly to aggregate shocks than sectoral shocks despite the fact that the true persistence in the model is the same to both shocks.