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

David Berger Ricardo Caballero Eduardo Engel
Northwestern MIT and NBER U. of Chile and NBER

March 19, 2013

Abstract

Conventional VAR procedures imply less persistence than there really is when adjustments underlying an aggregate variable are lumpy. This is relevant for non-, semi- and structural models in macroeconomics. The extent to which persistence is underestimated decreases with the level of aggregation, yet convergence is very slow and the bias is likely to be present for sectoral data in general and, in many cases, for aggregate data as well. Paradoxically, while idiosyncratic productivity and demand shocks smooth away microeconomic non-convexities and are often used to justify approximating aggregate dynamics with linear models, their presence exacerbates the bias. We propose procedures to correct for the bias and provide various applications. In one of them we find that the different speeds with which inflation responds to sectoral and aggregate shocks disappears once we correct for the missing aggregate dynamics.

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

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

1We are grateful to Filippo Altissimo, William Brainard, Xavier Gabaix, Pablo García, Robert Hall, Fabiano Schiaverdi, Harald Uhlig and seminar participants at Humboldt Universität, MIT, Universidad de Chile (CEA), University of Maryland, University of Pennsylvania, Yale University, NBER EFG Meeting, and the 2nd ECB/IMOP Workshop on Dynamic Macroeconomics, Hydra, for their comments on an earlier version of this paper. Financial support from NSF is gratefully acknowledged. This paper is an extensively revised version of “Adjustment is Much Slower than You Think,” NBER WP #9898.
1 Introduction

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

The main point of this paper is that when the microeconomic adjustment underlying an aggregate variable is lumpy, conventional VAR procedures often lead the researcher to conclude that there is less persistence than there really is. The extent to which persistence is underestimated decreases with the level of aggregation: linear models capture no persistence when applied to an individual series while the bias vanishes completely when they are applied to a series that aggregates infinitely many agents (Rotemberg, 1987). Interestingly, convergence is very slow: the bias is likely to be present in general for sectoral data and, quite often, for aggregate series as well. For example, even in the case of the U.S. Consumer Price Index, that aggregates approximately 70,000 prices, the bias turns out to be large, with the estimated half-life of shocks biased downward by approximately 40%.

We propose three procedures for correcting the bias we highlight in this paper. One estimates an ARMA specification for the aggregate of interest that captures the true underlying dynamics in the AR component. Another uses instrumental variables while the third approach uses a proxies for the underlying shocks. We also provide two detailed applications.

In the first application, we explain why estimates for the speed of adjustment of sectoral prices obtained using approaches tailored to the underlying lumpy behavior are much lower than those obtained with standard linear time-series models, thereby solving a puzzling result in Bils and Klenow (2004). Furthermore, we show that linear time series models deliver estimates in line with those obtained with nonlinear methods once the linear methods are applied correcting for the “missing-aggregate-persistence bias”.

Our second application revisits Boivin, Giannoni and Mihov’s (2009) finding that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks (see also Mackowiak, Moench and Wiederholt, 2008). In this case we show that once we correct for the missing-persistence bias the responses of inflation to both types of shocks look very similar.

The basic intuition underlying our results is obtained by comparing the true impulse response with that obtained using linear models, both for changes in the variable controlled by one agent (e.g., prices, capital, employment or durable stock). Even though in Section 5 we argue that the missing persistence bias is present quite generally, the underlying intuition is best understood assuming that shocks are i.i.d. so that an agent’s response every time it adjusts is equal to the sum of shocks that accumulated since the previous adjustment. We then have that the agent responds in
period \( t + k \) to a shock that took place in period \( t \) only if the agent adjusted in \( t \) and did not adjust in all periods between \( t \) and \( t + k - 1 \). It follows that the average response in \( t + k \) to a shock that took place in \( t \) is equal to the probability of having to wait exactly \( k \) periods until the first opportunity to adjust after the shock takes place. In the simple case where the arrival process that determines when adjustments take place follows a geometric distribution, as in the Calvo (1983) model, the nonlinear impulse response will be identical to that of an AR(1) process, with persistence parameter equal to the fraction of agents that do not adjust in a given period.

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

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

The remainder of the paper is organized as follows. Section 2 presents the Rotemberg (1987) equivalence result that justifies using linear time-series methods to estimate the dynamics for aggregates with lumpy microeconomic adjustment, as long as the number of units in the aggregate is infinite. Section 3 presents the missing-aggregate-persistence bias that arises when the number of units considered is finite. This section also provides results establishing the slow convergence to the Rotemberg limit. Section 4 describes various approaches to correct for the bias while Section 5 considers various extensions of the baseline model and shows that the bias is present, and con-
tinues being significant, in all of them. Section 6 studies two detailed applications and Section 7 concludes. Various appendices follow.

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

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

\[ a(L)\Delta y_t = \varepsilon_t, \quad (1) \]

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

The question that concerns us here is whether the estimated \( a(L) \) is likely to capture the true dynamics of the system when the underlying microeconomic variables are lumpy. Unless the effective number of underlying micro units is implausibly large, we will show that the answer often is ‘no’.

We set up the basic environment by constructing a simple model of microeconomic lumpy adjustment where the standard linear model in (1) is accurate when the effective number of microeconomic agents is infinitely large.

Let \( y_{it} \) denote the variable of concern at time \( t \) for agent \( i \)—e.g., a price, the level of employment, or its stock of capital—and \( y_{it}^* \) be the level the agent chooses if it adjusts in period \( t \). We will have that:

\[ \Delta y_{it} = \xi_{it}(y_{it}^* - y_{it-1}), \quad (2) \]

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

From a modeling perspective, discrete adjustment entails two basic features: First, periods of inaction are followed by abrupt adjustments to accumulated imbalances. Second, the likelihood of an adjustment increases with the size of the imbalance and is therefore state dependent. While the second feature is central for the macroeconomic implications of state-dependent models, it is not needed for the point we wish to raise in this paper. We therefore suppress it in this section and consider it when analyzing extensions in Section 5. That is, the special model we consider in this section corresponds to that in Calvo (1983), which is widely used in macroeconomic research. That is, we assume

\[ \Pr(\xi_{it} = 0) = \rho, \]
\[ \Pr(\xi_{it} = 1) = 1 - \rho. \quad (3) \]

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

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

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

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

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

That is, the effective number of units is equal to the inverse of the Herfindahl index. When all units contribute the same to the aggregate ($w_i = 1/n$) we have $N = n$, otherwise the effective number of units can be substantially lower than the actual number of units. We can now write the aggregate at time $t$, $y^N_t$, as:

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

Similarly we define the value of this aggregate should all units adjust, $y^{N*}_t$, as

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

### Technical Assumptions (Shocks)

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

1. The $v^A_t$'s are i.i.d. with mean $\mu_A$ and variance $\sigma^2_A > 0$.
2. The $v^I_{i,t}$'s are independent (across units, over time, and with respect to the $v^A$'s), identically distributed with zero mean and variance $\sigma^2_I > 0$.
3. The $\xi_{i,t}$'s are independent (across units, over time, and with respect to the $v^A$'s and $v^I$'s), identically distributed Bernoulli random variables with probability of success $\rho \in (0, 1]$.  

As Rotemberg (1987) showed, when $N$ goes to infinity, equation (4) for $\Delta y^\infty$ becomes:

$$\Delta y^\infty_t = (1 - \rho)(y^{\infty*}_t - y^{\infty}_{t-1}). \quad (5)$$
Taking first differences yields

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

which is the analog of Euler equations derived from a simple quadratic adjustment cost model applied to a representative agent.\(^2\)

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

Before doing so, let us close the loop by recovering equation (1) in this setup. For this, let us momentarily generalize the Technical Assumptions and assume that the process describing the average adjustment of units that adjust, \(\Delta y_t^{\infty*}\), is generated by:

\[ b(L) \Delta y_t^{\infty*} = \epsilon_t, \]

where the \(\epsilon_t\)'s are i.i.d. and \(b(L) \equiv 1 - \sum_{i=1}^q b_i L^i\) defines a stationary AR(q) for \(\Delta y_t^{\infty*}\). 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 \(\Delta y_t^*\) yields

\[ (1 - \rho L) b(L) \Delta y_t^\infty = (1 - \rho) \epsilon_t. \]

Comparing this expression with (1) we conclude that

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

Regardless of whether the researcher has any interest in recovering \(\rho\) per-se, the bias we highlight in this paper comes from a severe downward bias in its (implicit or explicit) estimate, 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 5.

### 3 The Missing Aggregate Persistence Bias

Of course, in reality the effective number of units, \(N\), in any aggregate is much smaller than infinity. The question that concerns us in this section is whether \(N\) is sufficiently large so that the limit result

\(^2\)The trend of using quadratic loss functions in economics was initiated by Holt et al. (1961) and continued by Tinsley (1971), Sims (1974) and Sargent (1978).
provides a good approximation.

Our main proposition states that the answer to this question depends on parameter values, in particular, on the relative importance of idiosyncratic to aggregate shocks and the degree of concentration. When both are high, the bias remains significant even at the economy-wide level. We argue that this is likely to be the case for various aggregates with lumpy microeconomic adjustment in the U.S. and, by extension, for smaller economies and sectoral data.

3.1 The Theory

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

**Proposition 1 (Aggregate Bias)**

Let \( \hat{\rho} \) denote the OLS estimator of \( \rho \) in

\[
\Delta y^N_t = \text{const.} + \rho \Delta y^N_{t-1} + \varepsilon_t. \tag{7}
\]

Let \( T \) denote the time series length. Then, under the Technical Assumptions, \( \lim_{T \to \infty} \hat{\rho} \) depends on the weights \( w_i \) only through \( N \) and

\[
\lim_{T \to \infty} \hat{\rho} \equiv \frac{K}{1 + K}, \quad \text{with} \quad K \equiv \frac{1 - \rho}{1 + \rho} (N - 1) - \left( \frac{\mu_A}{\sigma_A} \right)^2 \frac{1}{1 + \left( \frac{\sigma_I}{\sigma_A} \right)^2 + \left( \frac{\mu_A}{\sigma_A} \right)^2}. \tag{9}
\]

It follows that:

\[
\lim_{N \to \infty} \lim_{T \to \infty} \hat{\rho}^N = \rho. \tag{10}
\]

**Proof** See Appendix A.

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

One key factor in determining the magnitude of the bias is the effective number of units in the aggregate being considered. The bias is smaller when \( N \) is larger. Another key determinant of the speed of convergence of the aggregate with lumpy adjustment to its linear time-series version is the relative importance of idiosyncratic and aggregate shocks. When \( \sigma_I / \sigma_A \) is large, convergence is
slow and the bias remains significant at high levels of aggregation. Other factors that contribute to slow convergence is a larger drift (in absolute value) in the process driving the gap between desired and actual \( y \), and a larger amount of inertia as captured by the fraction of agents that do not adjust in any given period, \( \rho \).

### 3.2 The bias is large in practice

To put the relevance of this non-limit result in perspective, we consider next three examples where lumpy microeconomic adjustment has been well established: employment, prices, and investment. Table 1 reports how the estimated half-life of shocks varies for these aggregates with the effective number of units, \( N \).

#### Table 1: Slow Convergence

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>Frequency</th>
<th>100</th>
<th>400</th>
<th>1,000</th>
<th>4,000</th>
<th>10,000</th>
<th>40,000</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prices</td>
<td>monthly</td>
<td>0.257</td>
<td>0.464</td>
<td>0.767</td>
<td>1.744</td>
<td>2.699</td>
<td>3.886</td>
<td>4.595</td>
</tr>
<tr>
<td>Employment</td>
<td>quarterly</td>
<td>0.373</td>
<td>0.663</td>
<td>0.912</td>
<td>1.197</td>
<td>1.287</td>
<td>1.338</td>
<td>1.357</td>
</tr>
<tr>
<td>Investment</td>
<td>annual</td>
<td>0.179</td>
<td>0.356</td>
<td>\textbf{0.582}</td>
<td>1.333</td>
<td>2.167</td>
<td>3.397</td>
<td>4.265</td>
</tr>
</tbody>
</table>

Reported: half-life of shock inferred from estimation of (7), which is \(-\log_2/\log \hat{\rho}_\infty\) with \( \hat{\rho}_\infty = \text{plim}_{T \to \infty} \hat{\rho} \) obtained from Proposition 1. Parameters for prices: \( \rho = 0.86, \mu_A = 0.003, \sigma_A = 0.0054, \sigma_I = 0.048 \). Parameters for employment: \( \rho = 0.60, \mu_A = 0.005, \sigma_A = 0.03, \sigma_I = 0.25 \). Parameters for investment: \( \rho = 0.85, \mu_A = 0.12, \sigma_A = 0.056, \sigma_I = 0.50 \). Numbers in boldface correspond to the effective number of units for U.S. aggregates (CPI for prices, non-farm-business sector for employment and investment).

Let us begin with U.S. prices, a topic that has generated a new wave of research using the CPI micro database, following Bils and Klenow (2004). The results for prices, reported in the first row in Table 1, assume \( \rho = 0.86 \), in line with the median frequency of price adjustments for regular prices reported in Klenow and Kryvtsov (2008).\(^3\) Values for \( \mu_A \) and \( \sigma_A \) are taken from Bils and Klenow (2004), while \( \sigma_I \) is consistent with the value estimated in in Caballero et al (1997).\(^4\) The table shows

\(^3\)The average over the eight median frequencies reported by Nakamura and Steinsson for regular price changes suggest taking \( \rho = 0.89 \) which leads to a somewhat larger bias.

\(^4\)To go from the \( \sigma_I \) computed for employment in Caballero et al. (1997) to that of prices, we note that if the demand faced by a monopolistic competitive firm is isoelastic, its production function is Cobb-Douglas, and its capital fixed (which is nearly correct at high frequency), then (up to a constant):

\[
\hat{p}_{t} = (w_t - a_{t,L}) + (1 - \alpha_L)\hat{l}_{t}^{*}
\]

where \( \hat{p}_{t} \) and \( \hat{l}_{t}^{*} \) denote the logarithms of frictionless price and employment, \( w_t \) and \( a_{t,L} \) are the logarithm of the nominal wage and productivity, and \( \alpha_L \) is the labor share. It is straightforward to see that as long as the main source of idiosyncratic variance is demand, which we assume, \( \sigma_{I,t} = (1 - \alpha_L)\sigma_{l,t}^{*} \).
that the bias remains significant even for \( N = 10,000 \), which corresponds, approximately, to the effective number of prices used to calculate the CPI.\(^5\) In this case, the main reason for the stubborn bias is the high value of \( \sigma_I / \sigma_A \).

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

Finally, the third row in Table 1 reports the estimates for equipment investment, the most sluggish of the three series. The estimate of \( \rho, \mu_A \) and \( \sigma_A \), are from Caballero, Engel, and Haltiwanger (1995), and \( \sigma_I \) is consistent with that found in Caballero et al. (1997).\(^6\) Here the bias remains very large and significant throughout. In particular, when \( N = 986 \), which corresponds to the effective number of establishments for capital weights in the U.S. Non-Farm-Business sector in 2001, the estimated half-life of a shock is only 14% of the true half-life or, equivalently, the estimated frequency of adjustment, \( 1 - \rho \), is more than four times the true frequency. The reasons for this is the combination of a high \( \rho \), a high \( \mu_A \) (mostly due to depreciation), and a large \( \sigma_I \) (relative to \( \sigma_A \)).

Summing up, the missing persistence bias will be large at the sectoral level for inflation, employment and investment. Furthermore, linear time-series models will miss a substantial part of the dynamic behavior of U.S. inflation and investment at the aggregate level as well. The true half-life of a shock is close to twice its estimate for inflation and more than seven times its estimate for investment. Even though the setting we have used to gauge the magnitude of the bias is quite simple, in Section 5 we show that these conclusions are robust.

### 3.3 What is behind the bias and slow convergence?

Having established the proposition and the practical relevance of the bias, let us turn to the intuition behind the proof of the proposition. We do this in two steps. We first describe the genesis of the bias, which can be seen most clearly when \( N = 1 \). We then show why, for realistic parameter values, the extreme bias identified in the \( N = 1 \) case vanishes very slowly as \( N \) grows.

\(^5\)The median (mean) total number of observations per month between XXX and YYY is 66,582 (67,428). The median (mean) effective number of observations per month during this period is 10,328 (10,730).

\(^6\)To go from the \( \sigma_I \) computed for employment in Caballero et al. (1997) to that of capital, we note that if the demand faced by a monopolistic competitive firm is isoelastic and its production function is Cobb-Douglas, then \( \sigma_{I,b} = \sigma_{I,l} \).
3.3.1 The genesis of the bias

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

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

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

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

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

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

<table>
<thead>
<tr>
<th>Adjust in $t - 1$</th>
<th>Adjust in $t$</th>
<th>$\Delta y_{t-1}$</th>
<th>$\Delta y_t$</th>
<th>Contribution to $\text{Cov}(\Delta y_t, \Delta y_{t-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>$\Delta y_t \Delta y_{t-1} = 0$</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>0</td>
<td>$\Delta y_t^*$</td>
<td>$\Delta y_t \Delta y_{t-1} = 0$</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>$\sum_{k=0}^{l_t-1} \Delta y_{t-1-k}$</td>
<td>0</td>
<td>$\Delta y_t \Delta y_{t-1} = 0$</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>$\sum_{k=0}^{l_t-1} \Delta y_{t-1-k}$</td>
<td>$\Delta y_t^*$</td>
<td>$\text{Cov}(\Delta y_{t-1}, \Delta y_t) = 0$</td>
</tr>
</tbody>
</table>

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

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

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

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

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

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

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

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

This is the IRF for an AR(1) process obtained for aggregate inflation in the standard Calvo model (see, for example, Section 3.2 in Woodford, 2003).\(^7\)

What happened to Wold’s representation, according to which any stationary, purely non-deterministic, process admits an (eventually infinite) MA representation? Why do we obtain a Wold representation that is i.i.d., suggesting an infinitely fast response to shocks? What happens is that Wold’s representation only captures the first two moments of any process, it does not necessarily capture higher moments. In fact, Wold’s representation expresses the variable of interest as a distributed lag of innovations equal to the one-step-ahead linear forecast errors of $\Delta y$. Precisely because $E[\Delta y_t | \Delta y_{t+1}] = 0$, the one-step ahead forecast error of $\Delta y_{t+1}$ is equal to the variable being forecast, not to the innovation of economic interest, $\Delta y^*_t$. This misidentification will be present in any VAR model including variables with lumpy adjustment.

### 3.3.2 Slow convergence

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

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

---

\(^7\)As discussed in Caballero and Engel (2007), the impulse response for an individual unit and the corresponding aggregate will be the same for a broad class of macroeconomic models.
four different terms:

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

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

\[
\lim_{T \to \infty} \hat{\rho} = \frac{N \text{Cov}(\Delta y_{i,t}, \Delta y_{i,t-1}) + N(N-1) \text{Cov}(\Delta y_{i,t}, \Delta y_{j,t-1})}{N \text{Var}(\Delta y_{i,t}) + N(N-1) \text{Cov}(\Delta y_{i,t}, \Delta y_{j,t})}, \tag{13}
\]

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

### Table 3: Constructing the First Order Correlation

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

If \( N = 1 \), only the two within-agent terms remain, one in the numerator and one in the denominator. Since the covariance in the numerator is zero,\(^8\) \( \hat{\rho} \) is zero as well. This drag on \( \hat{\rho} \) remains present as \( N \) grows, but its relative importance declines since the between-agents covariances in the numerator and denominator are multiplied by terms of order \( N^2 \). This means that the reduction of the bias must come from the between-agents correlations at leads and lags, captured by the second expression in the numerator and denominator. The expression in the numerator is positive because not all individual units react to common shocks at the same time. The expression in the denominator is positive, because some do react at the same time. Either way it is clear that these expressions are proportional to the variance in aggregate, not idiosyncratic shocks. In fact, as summarized in the first row of Table 3:

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

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

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

However, as we mentioned above, these between-agents terms are proportional to the variance of aggregate shocks only. In contrast, the within-agent responsible for the biases are proportional

---

\(^8\)For simplicity we continue assuming \( \mu_A = 0 \).
to total uncertainty. In particular, the denominator of (13) is

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

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

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

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

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

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

### 4 Bias Corrections

This section studies three approaches to correct for the missing aggregate persistence bias. The first approach uses a proxy for target $y^*$, the second approach is based on an ARMA representation of
\( \Delta y_t^N \) while the third approach considers instrumental variables.

### 4.1 Using Proxies for the Target

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

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

(14)

with some proxy available for the regressor \( \Delta y^* \).

Equation (14) hints at a procedure for correcting the bias. Since the regressors are orthogonal, from Proposition 1 we have that the coefficient on \( \Delta y_{t-1} \) will be biased downward. By contrast, the true speed of adjustment can be estimated directly from the parameter estimate associated with \( \Delta y_t^* \), as long as the constraint that the sum of the coefficients on both regressors add up to one is not imposed. Of course, the estimate of \( \rho \) will be biased if the econometrician imposes the latter constraint. We summarize these results in the following proposition.

**Proposition 2 (Bias with Regressors)**

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

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

(15)

where \( \Delta y_t^* \) denotes the average shock in period \( t \), \( \sum w_i \Delta y_i^* \). Then, if (15) is estimated via OLS, and \( K \) defined in (9),

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

\[
\begin{align*}
\text{plim}_{T \to \infty} \hat{b}_0 &= \frac{K}{1 + K} \rho, \\
\text{plim}_{T \to \infty} \hat{b}_1 &= 1 - \rho;
\end{align*}
\]

(16) (17)

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

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

**Proof** See Appendix C.

Proposition 2 entails the general message that constructing a proxy for the target variable \( y^* \) can be very useful when estimating the dynamics of a macroeconomic variable with lumpy microeconomic adjustment. Also, it is important to avoid imposing constraints that hold only when \( N = \infty \). We apply this approach in Section 6.
4.2 ARMA Correction

The second correction we propose is based on a simple ARMA representation for $\Delta y_t^N$.

**Proposition 3 (ARMA Representation)**

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

$$\Delta y_t^N = \rho \Delta y_{t-1}^N + v_t - \theta v_{t-1},$$  \hspace{1cm} (18)

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

**Proof** See the Appendix. 

Equation (18) suggests a straightforward approach to estimating the adjustment speed parameter, $\rho$: Estimate an ARMA(1,1) process (18) and read off the estimate of $\rho$ from the AR-coefficient. That is, first estimate an ARMA model, next drop the MA polynomial and then make inferences about the implied dynamics using only the AR polynomial.

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

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

4.3 Instrumental Variables

Equation (18) in Proposition 1 suggests that lagged values of $\Delta y$ and $\Delta y^*$ (or components thereof) may be valid instruments to estimate $\rho$ in a regression of the form

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

More precisely, if $v_t = \Delta y_t^N$, then $\Delta y_{t-k}$ and $\Delta y_{t-k}^N$ 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, as it also includes $\Delta y_{t-1}^N$.

**Proposition 4 (Instrumental Variables)**
With the same notation and assumptions as in Proposition 1, we will have that $\Delta y_{t-k}^N$, $k \geq 2$ and $\Delta y_{t-j}^N$, $j \geq 1$ are valid instruments when estimating $\rho$ from

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

By contrast, $\Delta y_{t-1}^N$ is not a valid instrument.

Proof See Appendix D.

5 Extensions

Coming soon..

6 Applications

We now illustrate how the fast-slow bias manifests itself in two real world examples that involve price-setting. Pricing is an interesting context in which to study this bias because there is substantial evidence of infrequent adjustment at the individual good level. After showing that the fast-slow bias is substantial, we then show how to correct for the bias by using micro data to construct an empirical counterpart to the aggregate and sectoral shocks facing retail price-setters. Our first example shows that accounting for the fast-slow bias overturns Bils and Klenow’s rejection of the Calvo model from their now classic 2004 paper. We show that the fast-slow bias is empirically relevant in this example and that our bias correction procedure completely corrects for the bias. In our second example, we show that once the fast-slow bias is accounted for there is little evidence that sectoral prices respond faster to sectoral shocks than to aggregate results, contradicting the empirical results presented in Boivin, Gianonni and Mihov (2009) and Mackoviak, Moench and Wiederholt (2011).

6.1 Example 1: Bils-Klenow

In an intriguing contribution to their seminal paper, Bils and Klenow (2004) argue that one can use micro evidence on the frequency of adjustment at the sectoral level to directly test the Calvo model of price setting. Testing this assumption is important because the Calvo framework underlies many seminal papers in monetary economics. Their test of the Calvo mechanism is simple. In the Calvo framework with a frequency of adjustment $\lambda_s$, the dynamics of inflation for sector $s$ are given by:

$$\pi_{s,t} = (1 - \lambda_s)\pi_{s,t-1} + v_{s,t}$$

That is, the persistence of sectoral inflation rates is inversely related to the frequency of adjustment. Since Bils and Klenow (2004) had empirical measures of the frequency of adjustment by sector from the CPI micro data and sectoral inflation rate series from the BLS, they could easily test this central
prediction of the Calvo model. In particular, Bils and Klenow (2004) first estimated the persistence of sectoral inflation by estimating the following AR(1) regression:

\[ \pi_{s,t} = \beta_s \pi_{s,t-1} + e_{s,t} \] (19)

for 123 sectors. The Calvo model predicts that \( \hat{\beta}_s = 1 - \lambda_s \) but instead Bils and Klenow found that \( \hat{\beta}_s \ll 1 - \lambda_s \) for all 123 sectors in their data set (see Figures 3 and 4 of Bils and Klenow (2004)). They interpreted this evidence as a strong rejection of the Calvo mechanism. However, could this instead be a case of missing dynamics? Notice that this it is a textbook example of when the fast-slow bias will have bite: the underlying price data at the sectoral level have few observations and price adjustment is infrequent. Our theory implies that using a linear approximation when estimating the speeds of adjustment when they underlying variable exhibits lumpy adjustment leads to estimated responses much faster than true response, which is exactly what Bils-Klenow found.

To test this whether the fast-slow bias could be at work, we first want to get a back of the envelope estimate of whether the magnitude of the bias is quantitatively relevant in the above example. Towards that end, we calibrate a realistic multi-sector version Calvo model and compare the true adjustment frequencies with those estimated by linear time-series methods using simulated data. Our multi-sector model provides us with a simple laboratory to test whether the fast slow bias matters in this case.

<table>
<thead>
<tr>
<th>Table 5: Calibration Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Frequency of adjustment</td>
</tr>
<tr>
<td>Fraction price changes &gt; 0</td>
</tr>
<tr>
<td>Average size of increases</td>
</tr>
<tr>
<td>Average size of decreases</td>
</tr>
<tr>
<td>Std of sectoral inflation</td>
</tr>
</tbody>
</table>

The details of the calibration exercise are as follows. We calibrate a 66 sector version of the Calvo pricing model.\(^9\) For each sector, we set the average sectoral inflation rate to what is observed in the CPI micro data. We choose the standard deviation of the sectoral inflation rate series, the persistence and standard deviation of the sectoral idiosyncratic shock series (assumed to be an AR(1) in logs) to match the following four moments: the average size of price increases and decreases, the fraction of price changes that are price increases and the standard deviation of the sectoral inflation rate. In the model, the number of firms in each sector is given by the median (across time) number of firms for that sector in the micro BLS data and each firm was simulated for 252 periods\(^10\), which is the number of periods in the underlying data.

\(^9\)We have chosen 66 sectors (rather than 123) because this number represents the best balance between being disaggregated enough and having enough price change observations per sector (a constraint imposed by the CPI micro data).

\(^10\)More specifically, we simulated each firm for 1000 periods and only kept the last 252 observations.
Table 5 below shows that calibration results where all results are medians across the 66 sectors. As Table 5 shows, our multi-sector Calvo model does a good job of matching the average of these moments across sectors (should I show some other measures of fit as well?) Figure ?? is our version of Figures 3 and 4 in Bils and Klenow. It shows the relationship between the serial correlation of sectoral inflation and the frequency of price change in that sector. This is shown by the blue circles in Figure XX where the size of the markers (the circles) is proportional to the size of the standard errors across 250 simulations. The prediction of the Calvo model is shown by the solid black line. Our results are entirely consistent with Bils and Klenow’s: the estimated persistence of sectoral inflation rates is much lower than is implied by the Calvo model. That is, the blue circles almost always lie below the black line (the Calvo prediction) just as Bils and Klenow found using the CPI micro data. This exercise provides suggestive evidence that the fast-slow bias is empirically relevant, for even in our Calvo laboratory, this central prediction of the Calvo model is rejected by our simulations.

Figure 1:

Now that we have established that the fast-slow bias is relevant, we show how to correct for the bias using micro price data. In Section 4, we discussed three possible methods for correcting for this bias. In simulations, the ARMA correction and the IV approach were too fragile, thus we decided to use micro data to construct an empirical estimate of the shock. Our first measure of
the sectoral shock, $v_{s,t}$, is the average price change of those firms in sector $s$ that adjusted in both periods $t-1$ and $t$. Intuitively, in the Calvo model, when firm's adjust they adjust fully to all the shocks they have faced since they last adjusted. When firms adjust in this and last period, they adjust only to this periods shock. Thus by averaging across many firms within a sector one gets an unbiased estimate of the aggregate shock for period $t$: $v_{s,t} \approx \Delta y_{s,t}^*$.

A practical drawback of this intuitive approach is that on average we will only have $n_{s,t}\lambda_2$ observations to estimate $v_{s,t}$ hence there can be many periods with no (or very few) observations in sectors with a low frequency of adjustment. In the CPI RDB, the median number of observations and frequency adjustment across sectors is 132 and 0.079, respectively so in practice this limitation is severe. Thus we use a slightly modified approach that uses information on all cohorts of price changes.

Concretely, if a firm $f$ in sector $s$ adjusted in periods $t$ after last adjusting in period $t-k$ then it’s price change is given by:

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

(20)

where $e_{fst}$ denotes the sum of idiosyncratic shocks between $t$ and $t-k$. We use this insight to recover the time series of sectoral shocks for each sector. Assume that there are $T_s$ periods and $n_s$ price changes in sector $s$. Let $Y$ be the $n_s\times1$ matrix of these price changes, $X$ is the $n_s\times T$ design matrix where each element is equal to 1 if the corresponding price change was reacting to a shock at time $t$, and zero otherwise. Finally, $B$ is a $T\times1$ matrix of the sectoral shocks, $v_s$. Under the Calvo assumption, the error terms satisfies standard orthogonality restrictions so we can estimate this equation by OLS. This approach has a similar flavor to how Case-Shiller use repeated sales to estimate housing price indices. A great virtue of this approach is that it only requires $n_s\lambda_s$ rather than $n_s\lambda_2^s$ to estimate, a much less demanding empirical requirement.

In Appendix X we show that this estimation procedure generates consistent estimates for the $v_{s,t}$ if the underlying model is Calvo. We also show that in generalized $S_s$ models, this methodology estimates an affine transformation of the underlying shocks as the number of price changes goes to infinity. In Appendix XX, we show that estimating Equation (2) with only observations where $k=1$ is equivalent to our first approach.

Our methodology for estimating shocks is conceptually similar to Bils, Klenow and Malin’s (2012) "reset price inflation" concept since we are infering information about the underlying shocks by using information contained in all price changes in a given period. In simulations we found that our measure was more efficient in multi-sector Calvo models and of similar efficiency for $S_s$ simulations. In Appendix Y we present empirical results showing that our conclusions are robust to using reset price inflation as our measure of the shock.

Following Section 4, we implement our test by including our measure of the sectoral shock
in Equation (1). In particular, we run the following regression:

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

(21)

Our results from Section 4 suggest that $\hat{\gamma}_s$ is an unbiased estimate of $\lambda_s$. We estimate this regression sector by sector using simulation data from our calibrated multi-sector Calvo model. The results for each sector are represented by red circles in Figure X where each red circle represents one "corrected" frequency, $\lambda_{c,s}$. Our procedure is a clear empirical success as all the corrected estimates now lie around the Calvo prediction denoted by the solid line.

Motivated by this success, we now implement this approach using micro data on prices from the BLS. We use the CPI research database which contains individual price observations for the thousands of non-shelter items underlying the CPI. Prices are collected monthly for all items only in New York, Los Angeles, and Chicago, and so we restrict our analysis to these cities to ensure the representativeness of our sample. The database contains thousands of individual "quote-lines" with price observations for many months. In our data set, an average month contains about 10,000 different quote-lines. Quote-lines are the highest level of disaggregation possible and correspond to an individual item at a particular outlet. An example of a quote-line collected in the research database is 16 oz bag of frozen corn at a particular Chicago outlet. These quote-lines are then classified into various product categories called "Entry Level Items" or ELIs.

The ELIs can then be grouped into several levels of more aggregated product categories finishing with eleven major expenditure groups: processed food, unprocessed food, household furnishings, apparel, transportation goods, recreation goods, other goods, utilities, vehicle fuel, travel, and services. We choose to work at the two-digit or "Expenditure class" level of aggregate because it represented the best balance between being disaggregated enough and having enough price changes per sector per month. Additionally, we chose these were all sectors for which we could have data for the entire sample period. In the end, we have information for 66 sectors.

Much of the recent literature has discussed the difference between sales, regular price changes and product substitutions. In our benchmark analysis, we focus on regular price changes, excluding sales and product substitutions. We use the series excluding sales and product substitutions as our benchmark for two reasons: 1) Eichenbaum, Jaimovich, and Rebelo (2012) and Kehoe and Midrigan (2012) argue that the behavior of sales is often significantly different from that of regular or reference prices and that regular prices are likely to be the important object of interest for aggregate dynamics.

---

11 The most representative sample would be to use all bimonthly observations, but then many price changes are potentially missing. Some items are sampled monthly outside of NY, LA and Chicago, but these items are not representative, so we restrict our monthly analysis to these three cities.

12 There was a large revision to the ELI classification system in 1997.

13 Our definition of sectors at the expenditure class level is somewhat different than the BLS's because we link the sectors across the 1997 ELI revision so that we can have data spanning the entire time of our data set. Other researchers (Bils and Klenow(2004), Nakamura and Steinsson (2008)) focused on the periods 1988-1997 and post-1998 separately due to this changes in the definition of ELI's.
Thus, we choose to exclude sales in our benchmark analysis. 2) Product substitutions require a judgement on what portion of a price change is due to quality adjustment and which component is a pure price change. Thus, this introduces measurement error in the calculation of price changes at the time of product substitution. Bils (2009) shows that these errors can be substantial. For this reason, we exclude product substitutions from our benchmark analysis. Nevertheless, we have also repeated the analysis including product substitutions and found similar results.

As a first step we replicate Bils and Klenow’s (2004) results for our 66 sectors. First we estimate Equation (1) using the micro data. Denote the implied frequency estimates as $\lambda^{VAR}_s = 1 - \hat{\beta}_s$. Similar to Bils and Klenow (2004), we find that $\hat{\beta}_s \ll 1 - \lambda^{micro}_s$, where $\lambda^{micro}_s$ is true frequency of adjustment from the micro data. Next we estimate Equation (1) using our constructed shock measure. Denote the coefficient on our sectoral shock measure $\lambda^c_s$. To gauge the extent to which the $\lambda^c_s$ correct the missing-dynamics bias, we regress the change in estimated speed of adjustment we achieve in a given sector on the magnitude of the bias. That is, we estimate by OLS the following equation:

$$(\lambda^c_s - \lambda^{VAR}_s) = \alpha + \eta(\lambda^{micro}_s - \lambda^{VAR}_s)$$

Here $\eta$ is the coefficient of interest as it captures the extent to which our bias correction actually decreases the bias. A value of 0 suggests no correction and a value of 1 can be interpreted as a full correction.

Table X: Lambda Regressions

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

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table X shows the results. As you can see in both our multi-sector Calvo simulation and in the CPI database, our empirical is an unequivocal success. For the CPI, the estimated value of $\eta$ is not statistically different from one. This suggests that the conclusion drawn by Bils and Klenow (2004) that the micro data strongly reject the Calvo mechanism is not warranted, rather their results seemed to have been driven entirely by the fast-slow bias. More importantly, this example shows
the fast-slow bias is empirically relevant at the sectoral level and that an innovative use of micro
data can help one overcome this bias.

6.2 Example 2: Faster response of to sectoral rather than aggregate shocks?

The recent theoretical literature on sticky-information and costly observation models points
out that there is no reason why prices should adjust equally fast to different types of shocks. In
recent paper, Boivin, Gianonni and Mihov (2009) (henceforth, BGM) provides empirical evidence
that sectoral inflation responds much faster to sectoral shocks than to aggregate shocks, which is
consistent with both of these classes of theoretical models.

To understand BGM’s approach, we must first introduce some terminology. Define \( \Pi_t \) as a vec-
tor of sectoral inflation rates from the BEA and the PPI. They assume that \( \Pi_t \) can be decomposed
into the sum of a few common factors, \( C_t \), and a sectoral component, \( e_t \). Here \( \Lambda \) is a matrix of factor
loadings for the common factor. These factor loadings are allowed to differ across sectors.

\[
\Pi_t = \Lambda C_t + e_t
\]  

(22)

BGM extract principal components from the large data set \( \Pi_t \) to obtain consistent estimates of
the common factors.\(^{14}\) Next, they regress each sectoral inflation series on the common factors,
denoting the predicted component the aggregate component, \( \pi_{st}^{agg} \), and denoting the residual the
sector specific component, \( \pi_{st}^{sect} \).

\[
\pi_{it} = \lambda_i' C_t + \pi_{it}^{agg} + \pi_{it}^{sect}
\]  

(23)

This formulation allows one to disentangle the fluctuations in sectoral inflation rates due to the
macroeconomic factors— represented here by the common components \( C_t \) which have a diffuse
effect on all data series— from those due to sector-specific conditions represented by the term \( e_{it} \).

To calculate IRFs with respect to the common and sectoral shocks, BGM fit separate AR(13)s to \( \pi_{st}^{agg} \)
and \( \pi_{st}^{sect} \).

\[
x_t = \sum_{k=1}^{13} a_k x_{t-k} + e_t.
\]

BGM measure the persistence of shocks persistence by: \( \sum_{k=1}^{13} a_k \). Under the Calvo assumption, this
sum equals to \( 1 - \lambda \). To start, we reproduce their benchmark results. The results from are shown in
Table YYY:

\(^{14}\)Stock and Watson (2002) show that the principal components consistently recover the space spanned by the factors
when the number of series is large and the number of principal components used is at least as large as the true number
of factors.
Persistence measure

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

As Table YYY shows, both the mean and median persistence of the "aggregate" component is much larger than the mean and median persistence of the "sectoral" component. This is the main evidence that leads BGM to conclude that sectoral inflation rates respond much more quickly to sectoral shocks than to aggregate shocks in the U.S. A subsequent paper by Mackoviak, Moench and Wiederholt (2011), using a different methodology found quantitatively similar results using the CPI data. Both papers conclude that this difference is persistence is strong evidence in favor of sticky-information models. However, notice that BGM measure persistence of each component by regressing each component on lags of itself. Since the underlying prices adjust infrequently and there are not many prices underlying these sectoral inflation series, could BGM’s results be driven by the fast-slow bias?

To investigate this hypothesis, we use the same shock measures that we computed from CPI micro data that were discussed in depth in the previous section. That is, we have data for 66 sectoral inflation series from the CPI for the period 1988:03-2007:12. Define $V_t$ as a be a vector with our period $t$ sectoral shock measure for all sectors. Our proxy for the aggregate shock is the first $K$ principal components of $V$:

$$m^k_t, \quad k = 1, 2, ..., K.$$  

We also allow for distributed lags (the $j’s$) in relation between sectoral innovations and aggregate innovations:

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

however our results are robust to ignoring these distributed lags. The end result is $K$ aggregate shocks, $m^k_t$, and a sectoral shock, $x_{st}$, for each of 66 sectors from the CPI. Once we have these shocks, we estimate:

$$\pi_{st} = \sum_{k=1}^{K} \sum_{j\geq 0} \eta^k_{sj} m_{t-j}^k + \sum_{j\geq 0} \nu_{sj} x_{s,t-j}.$$  

Our persistence measure is slightly different than BGM’s (though the results are robust to using their
persistence measure). The expected response times to each type of shocks in the following manner:

\[
\tau^{\text{sec}}_s = \sum_{j \geq 0} j \nu^k_{s_j} / \sum_{j \geq 0} \nu^k_{s_j},
\]

\[
\tau^{\text{agg}, k}_s = \sum_{j \geq 0} j \eta^k_{s_j} / \sum_{j \geq 0} \eta^k_{s_j},
\]

\[
\tau^{\text{agg}}_s = \text{median}_k \tau_{s,k}.
\]

Figure 2:

<table>
<thead>
<tr>
<th>PCs</th>
<th>nlags</th>
<th>BGM agg</th>
<th>BGM sect</th>
<th>First ν agg</th>
<th>First ν sect</th>
<th>Second ν agg</th>
<th>Second ν sect</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>1.79</td>
<td>0.87</td>
<td>5.58</td>
<td>5.62</td>
<td>6.31</td>
<td>5.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.32)</td>
<td>(0.24)</td>
<td>(0.36)</td>
<td>(0.39)</td>
<td>(0.51)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>1.37</td>
<td>1.10</td>
<td>5.91</td>
<td>5.93</td>
<td>6.35</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.45)</td>
<td>(0.25)</td>
<td>(0.38)</td>
<td>(0.60)</td>
<td>(0.34)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>1.70</td>
<td>0.96</td>
<td>5.99</td>
<td>5.31</td>
<td>5.99</td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(0.26)</td>
<td>(0.34)</td>
<td>(0.45)</td>
<td>(0.29)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>1.50</td>
<td>0.76</td>
<td>5.92</td>
<td>6.19</td>
<td>6.16</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.34)</td>
<td>(0.23)</td>
<td>(0.26)</td>
<td>(0.50)</td>
<td>(0.20)</td>
<td>(0.51)</td>
</tr>
</tbody>
</table>

Notice that since we have a direct proxy for both shocks, our measures of persistence to these shocks are not susceptible to the fast-slow critique. The results are shown in the above table. All results are medians across sectors and the standardized interquartile ranges are shown in parentheses. The first two columns show the results from replicating the BGM approach using the CPI data. As you can see, in all specifications the persistence of the aggregate shock is significantly greater than the persistence of the sectoral shock, just as BGM asserted. The next two columns show the results from using our approach. Two facts stand out. First, the estimate response time of both the aggregate and sectoral component increases relative to BGM’s approach. This is strong evidence in favor of the empirical relevance of the fast-slow bias. Furthermore, the magnitude of the increase is greater for the sectoral shock. This is also consistent with the empirical relevance of the
fast-slow bias since this bias should be more severe for the sectoral component and the aggregate component. Second, after correcting for the fast-slow bias, there is no significant difference between the estimated response times of sectoral inflation series to aggregate and sectoral shocks. In other words, there is no longer any evidence that firms respond differently to aggregate and sectoral shocks.

7 Conclusion

To reiterate, the researcher may not give a damn about lambda. The issue is not this. It is about estimating IRFs and aggregate dynamics for ANY purpose, including calibration!

Note that idiosyncratic shocks play a dual role. On one hand, they smooth the micro-nonconvexities as we aggregate until they "eye" can't see them. On the other, they ensure that bias remains large well beyond that smoothing point.

Distinct from Golosov-Lucas effect. We are working on it.
References


APPENDIX

A Proof of Propositions

Proof of Proposition 1

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

\[
x_{i,t} \equiv y_{i,t}^* - y_{i,t-1}.
\]

Give the Technical Assumptions, we have that \( x_{i,t} \) equals the unit’s accumulated shocks since it last adjusted.

The following expressions characterize the dynamics of \( x_{i,t} \) as well as relating this variable to changes in the variable of interest:

\[
x_{i,t+1} = (1 - \xi_{i,t})x_{i,t} + \Delta y_{i,t+1}^*, \quad (24)
\]

\[
\Delta y_{i,t} = \xi_{i,t} x_{i,t}. \quad (25)
\]

In what follows, subindices \( i \) and \( j \) denote different units.

We first derive the following unconditional expectations:

\[
E[x_{i,t}] = \frac{\mu_A}{1 - \rho}, \quad (26)
\]

\[
E[\Delta y_{i,t}] = \mu_A, \quad (27)
\]

\[
E[\Delta y_{N,t}] = \mu_A, \quad (28)
\]

\[
E[x_{i,t} x_{j,t}] = \frac{1}{1 - \rho^2} \left[ \sigma_A^2 + \frac{1 + \rho}{1 - \rho} \mu_A^2 \right]. \quad (29)
\]

\[
E[x_{i,t}^2] = \frac{1}{1 - \rho} \left[ \sigma_A^2 + \sigma_l^2 + \frac{1 + \rho}{1 - \rho} \mu_A^2 \right]. \quad (30)
\]

From (24) and the Technical Assumption in the main text we have:

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

The above expression leads to (26) once we note that the stationarity of \( x_{i,t} \) implies \( E[x_{i,t+1}] = E[x_{i,t}] \).

Equation (27) follows from (26) and Technical Assumption 3. Equation (28) follows directly from (27).
To derive (29), we note that, from (24)

\[ E[x_{i,t+1}y_{j,t+1}] = E[(1 - \xi_{i,t})x_{i,t} + \Delta y^*_{i,t+1}][(1 - \xi_{j,t})x_{j,t} + \Delta y^*_{j,t+1}] \]

\[ = E[(1 - \xi_{i,t})x_{i,t}(1 - \xi_{j,t})x_{j,t}] + E[\Delta y^*_{i,t+1}(1 - \xi_{j,t})x_{j,t}] + E[(1 - \xi_{i,t})x_{i,t}\Delta y^*_{j,t+1}] + E[\Delta y^*_{i,t+1}\Delta y^*_{j,t+1}] \]

\[ = \rho^2 E[x_{i,t}x_{j,t}] + 2 - \rho \mu_A^2 + (\mu_A^2 + \sigma_A^2), \]

where we used the Technical Assumptions, (26) and \( i \neq j \). Noting that \( x_{i,t}x_{j,t} \) is stationary and therefore \( E[x_{i,t}x_{j,t}] = E[x_{i,t-1}x_{j,t-1}] \), the above expression leads to (29).

Finally, to prove (30), we note that, from (24) we have

\[ E[x_{i,t}^2] = E[(1 - \xi_{i,t})x_{i,t}^2] + 2E(1 - \xi_{i,t})x_{i,t}\Delta y^*_{i,t+1} + E[(\Delta y^*_{i,t+1})^2] \]

\[ = \rho E[x_{i,t}^2] + 2 - \rho \mu_A^2 + (\sigma^2 + \sigma_A^2 + \mu_A^2), \]

where we used that \( (1 - \xi_{i,t})^2 = 1 - \xi_{i,t} \), (26) and the Technical Assumptions. Stationarity of \( x_{i,t} \) (and therefore \( x_{i,t}^2 \)) and some simple algebra complete the proof.

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

We have:

\[ \text{Cov}(\Delta y_{i,t+1}, \Delta y_{j,t}) = E[\Delta y_{i,t+1}\Delta y_{j,t}] - \mu_A^2 = E[\xi_{i,t+1}x_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)E[x_{i,t+1}\xi_{i,t}x_{j,t}] - \mu_A^2 \]

\[ = (1 - \rho)E[\xi_{i,t+1}x_{i,t} + \Delta y^*_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)E[(1 - \xi_{j,t})x_{i,t}x_{j,t}] + (1 - \rho)E[\Delta y^*_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 \]

\[ = (1 - \rho)E[(1 - \xi_{j,t})x_{j,t}] + (1 - \rho)E[\Delta y^*_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 \]

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

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

\[ \text{Cov}(\Delta y_{i,t+1}, \Delta y_{j,t}) = E[\xi_{i,t+1}x_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)E[x_{i,t+1}\xi_{i,t}x_{j,t}] - \mu_A^2 \]

\[ = (1 - \rho)E[(1 - \xi_{i,t})x_{i,t} + \Delta y^*_{i,t+1}\xi_{j,t}x_{j,t}] - \mu_A^2 = \rho(1 - \rho)E[x_{i,t}x_{j,t}] + (1 - \rho)\mu_A^2 - \mu_A^2 = \frac{1}{\rho + 1} \rho \sigma_A^2. \]

\[ \text{Cov}(\Delta y_{i,t}, \Delta y_{j,t}) = E[\xi_{i,t}x_{i,t}\xi_{j,t}x_{j,t}] - \mu_A^2 = (1 - \rho)^2E[x_{i,t}x_{j,t}] - \mu_A^2 = \frac{1}{\rho + 1} \sigma_A^2. \]

Finally, the variance term is obtained as follows:

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

**Proof of Proposition 2**
Part (i) follows trivially from Proposition 1 and the fact that both regressors are uncorrelated. To prove (ii) we first note that:

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

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

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

Both covariances on the r.h.s. are calculated using (24), yielding \(\sigma_A^2 + \sigma_T^2\) and \(\sigma_A^2\), respectively. Expressions for \(\text{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)\) and \(\text{Var}(\Delta y_t^N)\) are obtained using an analogous decomposition and the covariances and variances from Table 3. We have all the terms for the expression above for \(\hat{b}_1\), the remainder of the proof is some tedious but otherwise straightforward algebra.

**Proof of Proposition 3**

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

\[
\gamma_k = \rho \gamma_{k-1}, \quad k \geq 2. \tag{31}
\]

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

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

Proving that \(\theta\) tends to zero as \(N\) tends to infinity is straightforward.

We have:

\[
\begin{align*}
E[\Delta y_{t+k}^N & \Delta y_t^N] = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{i,t+k} x_{i,t+k} \xi_{j,t} x_{j,t}] \\
& = (1 - \rho) \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[x_{i,t+k} \xi_{j,t} x_{j,t}] \\
& = (1 - \rho) \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[(1 - \zeta_{i,t+k-1}) x_{i,t+k-1} + \Delta y_{i,t+k}^* \xi_{j,t} x_{j,t}] \\
& = (1 - \rho) \rho \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[x_{i,t+k} \xi_{j,t+k-1} x_{j,t}] + (1 - \rho) \mu_A \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{j,t} x_{j,t}] \\
& = \rho \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j E[\xi_{j,t+k-1} x_{i,t+k-1} \xi_{j,t} x_{j,t}] + (1 - \rho) \mu_A^2 \\
& = \rho E[\Delta y_{t+k-1}^N \Delta y_t^N] + (1 - \rho) \mu_A^2.
\end{align*}
\]

\[15\]Here we are using Theorem 1 in Engel (1984) characterizing ARMA processes in terms of difference equations satisfied by their autocorrelation function.
where in the fourth step we assumed \( k \geq 2 \), since we used that \( \xi_{i,t+k-1} \) and \( \xi_{j,t} \) are independent even when \( i = j \). Noting that \( y_k = (E[\Delta y_{t+k}^N \Delta y_t^N] - \mu_2^*)/\text{Var}(\Delta y_t) \) and using the above identity yields (31) and concludes the proof.

**Proof of Proposition 4**

We have:

\[
\Delta y_t^N = \sum_{i} w_i \xi_{i,t} x_{i,t} = \sum_{i} w_i \xi_{i,t} (y_{i,t}^* - y_{i,t-1}) = \sum_{i} w_i (1-\rho)(y_{i,t}^* - y_{i,t-1}) + \sum_{i} w_i (\xi_{i,t-1} + \rho)(y_{i,t}^* - y_{i,t-1}).
\]

Similarly

\[
\Delta y_{t-1}^N = \sum_{i} w_i (1-\rho)(y_{i,t-1}^* - y_{i,t-2}) + \sum_{i} w_i (\xi_{i,t-1} - 1 + \rho)(y_{i,t-1}^* - y_{i,t-2}).
\]

Subtracting the latter from the former and rearranging terms yields

\[
\Delta y_t^N = \rho \Delta y_{t-1}^N + (1-\rho)\Delta y_t^* + \epsilon_t^N \tag{32}
\]

with

\[
\epsilon_t^N = \sum_{i} w_i \left[ (\xi_{i,t-1} + \rho)(y_{i,t}^* - y_{i,t-1}) - (\xi_{i,t-1} - 1 + \rho)(y_{i,t-1}^* - y_{i,t-2}) \right]. \tag{33}
\]

The extra term \( \epsilon_t^N \) on the r.h.s. of (33) explains why \( \Delta y_{t-1}^N \) is not a valid instrument: \( \Delta y_{t-1}^N \) is correlated with \( \epsilon_t^N \) because both include \( \xi_{i,t-1} \) terms. Of course, \( \epsilon_t^N \) tends to zero as \( N \) tends to infinity: its mean is zero and a calculation using many of the expressions derived in the proof of Proposition 1 shows that

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

It follows from (32), (33) and Technical Assumption 3 that \( \epsilon_t \) is uncorrelated with \( \Delta y_t^s \), for all \( s \), which implies that \( \Delta y_{t-s}^* \) is a valid instrument for \( s \geq 1 \). And since \( \Delta y_{i,t-k} \) are uncorrelated with \( \xi_{i,t} \) and \( \xi_{i,t-1} \) for \( k \geq 2 \), we have that lagged values of \( \Delta y \), with at least two lags, are valid instruments as well. 

\[
\]