Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Adjustment Is Much Slower Than You Think

Ricardo J. Caballero
Eduardo M.R.A. Engel

Working Paper 03-25
July 21, 2003

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Adjustment is Much Slower than You Think

Ricardo J. Caballero Eduardo M.R.A. Engel*

July 21, 2003

Abstract

In most instances, the dynamic response of monetary and other policies to shocks is infrequent and lumpy. The same holds for the microeconomic response of some of the most important economic variables, such as investment, labor demand, and prices. We show that the standard practice of estimating the speed of adjustment of such variables with partial-adjustment ARMA procedures substantially overestimates this speed. For example, for the target federal funds rate, we find that the actual response to shocks is less than half as fast as the estimated response. For investment, labor demand and prices, the speed of adjustment inferred from aggregates of a small number of agents is likely to be close to instantaneous. While aggregating across microeconomic units reduces the bias (the limit of which is illustrated by Rotemberg's widely used linear aggregate characterization of Calvo's model of sticky prices), in some instances convergence is extremely slow. For example, even after aggregating investment across *all* establishments in U.S. manufacturing, the estimate of its speed of adjustment to shocks is biased upward by more than 80 percent. While the bias is not as extreme for labor demand and prices, it still remains significant at high levels of aggregation. Because the bias rises with disaggregation, findings of microeconomic adjustment that is substantially faster than aggregate adjustment are generally suspect.

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

Keywords: Speed of adjustment, discrete adjustment, lumpy adjustment, aggregation, Calvo model, ARMA process, partial adjustment, expected response time, monetary policy, investment, labor demand, sticky prices, idiosyncratic shocks, impulse response function, Wold representation, time-to-build.

*Respectively: MIT and NBER; Yale University and NBER. We are grateful to William Brainard, Pablo García, Harald Uhlig and seminar participants at Humboldt Universität, Universidad de Chile (CEA) and University of Pennsylvania for their comments. Financial support from NSF is gratefully acknowledged.

1 Introduction

The measurement of the dynamic response of economic and policy variables to shocks is of central importance in macroeconomics. Usually, this response is estimated by recovering the speed at which the variable of interest adjusts from a linear time-series model. In this paper we argue that, in many instances, this procedure significantly underestimates the sluggishness of actual adjustment.

The severity of this bias depends on how infrequent and lumpy the adjustment of the underlying variable is. In the case of single policy variables, such as the federal funds rate, or individual microeconomic variables, such as firm level investment, this bias can be extreme. If no source of persistence other than the discrete adjustment exists, we show that regardless of how sluggish adjustment may be, the econometrician estimating linear autoregressive processes (partial adjustment models) will erroneously conclude that adjustment is instantaneous.

Aggregation across establishments reduces the bias, so we have the somewhat unusual situation where estimates of a microeconomic parameter using aggregate data are less biased than those based upon microeconomic data. Lumpiness combined with linear estimation procedures is likely to give the false impression that microeconomic adjustments is significantly faster than aggregate adjustment.

We also show that, when aggregating across units, convergence to the correct estimate of the speed of adjustment is extremely slow. For example, for U.S. manufacturing investment, even after aggregating across all the continuous establishments in the LRD (approximately 10,000 establishments), estimates of speed of adjustment are 80 percent higher than actual speed — at the two-digit level the bias easily can exceed 400 percent. Similarly, estimating a Calvo (1983) model using standard partial-adjustment techniques is likely to underestimate the sluggishness of price adjustments severely. This is consistent with the recent findings of Bils and Klenow (2002), who report much slower speeds of adjustment when looking at individual price adjustment frequencies than when estimating the speed of adjustment with linear time-series models. We also show that estimates of employment adjustment speed experience similar biases.

The basic intuition underlying our main results is the following: In linear models, the estimated speed of adjustment is inversely related to the degree of persistence in the data. That is, a larger first order correlation is associated with lower adjustment speed. Yet this correlation is always zero for an individual series that is adjusted discretely (and has i.i.d. shocks), so that the researcher will conclude, incorrectly, that adjustment is infinitely fast. To see that this crucial correlation is zero, first note that the product of current and lagged changes in the variable of concern is zero when there is no adjustment in either the current or the preceding period. This means that any non-zero serial correlation must come from realizations in which the unit adjusts in two consecutive periods. But when the unit adjusts in two consecutive periods, and whenever it acts it catches up with all accumulated shocks since it last adjusted, it must be that the later adjustment only involves the latest shock, which is independent from the shocks included in the previous adjustment.

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 us to recover the microeconomic speed of adjustment. The larger this common component is —as measured, for example, by the variance of aggregate shocks relative to the variance of idiosyncratic shocks— the faster the estimate converges to its true value as the number of agents grows. In practice, the variance of aggregate shocks is significantly smaller than that of idiosyncratic shocks, and convergence takes place at a very slow pace.

In Section 2 we study the bias for microeconomic and single-policy variables, and illustrate its importance when estimating the speed of adjustment of monetary policy. Section 3 presents our aggregation results and highlights slow convergence. We show the relevance of this phenomenon for parameters consistent with those of investment, labor, and price adjustments in the U.S. Section 4 discusses partial solutions and extensions. It first extends our results to dynamic equations with contemporaneous regressors, such as those used in price-wage equations, or output-gap inflation models. It then illustrates an ARMA method to reduce the extent of the bias. Section 5 concludes and is followed by an appendix with technical details.

2 Microeconomic and Single-Policy Series

When the task of a researcher is to estimate the speed of adjustment of a state variable —or the implicit adjustment costs in a quadratic adjustment cost model (see e.g., Sargent 1978, Rotemberg 1987)— the standard procedure reduces to estimating variations of the celebrated *partial adjustment model* (PAM):

$$\Delta y_t = \lambda(y_t^* - y_{t-1}), \tag{1}$$

where y and y^* represent the actual and optimal levels of the variable under consideration (e.g., prices, employment, or capital), and λ is a parameter that captures the extent to which imbalances are remedied in each period. Taking first differences and rearranging terms leads to the best known form of PAM:

$$\Delta y_t = (1-\lambda)\Delta y_{t-1} + v_t, \tag{2}$$

with $v_t \equiv \lambda \Delta y_t^*$.

In this model, λ is thought of as the speed of adjustment, while the expected time until adjustment (defined formally in Section 2) is $(1-\lambda)/\lambda$. Thus, as λ converges to one, adjustment occurs instantaneously, while as λ decreases, adjustment slows down.

Most people understand that this model is only meant to capture the first-order dynamics of more realistic but complicated adjustment models. Perhaps most prominent among the latter, many microeconomic variables exhibit only infrequent adjustment to their optimal level (possibly due to the presence of fixed costs of adjustments). And the same is true of policy variables, such as the federal funds rate set by the monetary authority in response to changes in aggregate conditions. In what follows, we inquire how good the estimates of the speed of adjustment from the standard partial adjustment approximation (2) are, when actual adjustment is discrete.

2.1 A Simple Lumpy Adjustment Model

Let y_t denote the variable of concern at time t —e.g., the federal funds rate, a price, employment, or capital— and y_t^* be its optimal counterpart. We can characterize the behavior of an individual agent in terms of the equation:

$$\Delta y_t = \xi_t (y_t^* - y_{t-1}), \tag{3}$$

where ξ_t satisfies:

$$\begin{aligned} \Pr\{\xi_t = 1\} &= \lambda, \\ \Pr\{\xi_t = 0\} &= 1 - \lambda. \end{aligned} \tag{4}$$

From a modelling perspective, discrete adjustment entails two basic features: (i) periods of inaction followed by abrupt adjustments to accumulated imbalances, and (ii) increased likelihood of an adjustment with the size of the imbalance (state dependence). 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. Therefore, we suppress it.[1]

It follows from (4) that the *expected* value of ξ_t is λ. When ξ_t is zero, the agent experiences inaction; when its value is one, the unit adjusts so as to eliminate the accumulated imbalance. We assume that ξ_t is independent of $(y_t^* - y_{t-1})$ (this is the simplification that Calvo (1983) makes vis-a-vis more realistic state dependent models) and therefore have:

$$\mathrm{E}[\Delta y_t \mid y_t^*, y_{t-1}] = \lambda(y_t^* - y_{t-1}), \tag{5}$$

which is the analog of (1). Hence λ represents the *adjustment speed* parameter to be recovered.

2.2 The Main Result: (Biased) Instantaneous Adjustment

The question now arises as to whether, by analogy to the derivation from (1) to (2), the standard procedure of estimating

$$\Delta y_t = (1 - \lambda)\Delta y_{t-1} + \varepsilon_t, \tag{6}$$

recovers the *average* adjustment speed, λ, when adjustment is lumpy. The next proposition states that the answer to this question is clearly no.

Proposition 1 (Instantaneous Estimate)

Let $\hat{\lambda}$ denote the OLS estimator of λ in equation (6). Let the Δy_t^*'s be i.i.d. with mean 0 *and variance* σ^2,

[1] The special model we consider —i.e., without feature (ii)— is due to Calvo (1983) and was extended by Rotemberg (1987) to show that, with a continuum of agents, *aggregate* dynamics are indistinguishable from those of a representative agent facing quadratic adjustment costs. One of our contributions is to go over the aggregation steps in more detail, and show the problems that arise before convergence is achieved

and let T denote the time series length. Then, regardless of the value of λ:

$$\text{plim}_{T\to\infty}\hat{\lambda} = 1. \tag{7}$$

Proof See Appendix B.1. ∎

While the formal proof can be found in the appendix, it is instructive to develop its intuition in the main text. If adjustment were smooth instead of lumpy, we would have the classical partial adjustment model, so that the first order autocorrelation of observed adjustments is $1-\lambda$, thereby revealing the speed with which units adjust. But when adjustment is lumpy, the correlation between this period's and the previous period's adjustment necessarily is zero, so that the implied speed of adjustment is one, independent of the true value of λ. To see why this is so, consider the covariance of Δy_t and Δy_{t-1}, noting that, because adjustment is complete whenever it occurs, we may re-write (3) as:

$$\Delta y_t = \xi_t \sum_{k=0}^{l_t-1} \Delta y^*_{t-k} = \begin{cases} \sum_{k=0}^{l_t-1} \Delta y^*_{t-k} & \text{if } \xi_t = 1, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where l_t denotes the number of periods since the last adjustment took place, (as of period t).[2]

Table 1: CONSTRUCTING THE MAIN COVARIANCE

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

There are four scenarios to consider when constructing the key covariance (see Table 1): If there was 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 disjoint 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.

[2]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.

2.3 Robust Bias: Infrequent and Gradual Adjustment

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

Let us modify our basic model so that equation (3) 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:

$$\Delta y_t = \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + (1 - \sum_{k=1}^{K} \phi_k)\Delta\tilde{y}_t. \tag{9}$$

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 he adjusts, he does so only gradually.

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

$$\Delta\tilde{y}_t = (1-\lambda)\Delta\tilde{y}_{t-1} + v_t. \tag{10}$$

Replacing (10) into (9), yields the following linear equation in terms of the observable, Δy_t:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + \varepsilon_t, \tag{11}$$

with

$$\begin{aligned} a_1 &= \phi_1 + 1 - \lambda, \\ a_k &= \phi_k - (1-\lambda)\phi_{k-1}, \qquad k = 2, \ldots, K, \\ a_{K+1} &= -(1-\lambda)\phi_K, \end{aligned} \tag{12}$$

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

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

Proposition 2 (Omitted Source of Sluggishness)

Let all the assumptions in Proposition 1 hold, with $\tilde{y}$ in the role of y. Also assume that (9) applies, with all roots of the polynomial $1 - \sum_{k=1}^{K} \phi_k z^k$ outside the unit disk. Let $\hat{a}_k, k = 1, \ldots, K+1$ denote the OLS estimates of equation (11).

Then:

$$\begin{aligned} \text{plim}_{T\longrightarrow\infty}\hat{a}_k &= \phi_k, \qquad k=1,\ldots,K, \\ \text{plim}_{T\longrightarrow\infty}\hat{a}_{K+1} &= 0. \end{aligned} \tag{13}$$

Proof See Appendix B.1. ∎

Comparing (12) and (13) we see that the proposition simply reflects the fact that the (implicit) estimate of λ is one.

The mapping from the biased estimates of the a_k's to the speed of adjustment is slightly more cumbersome, but the conclusion is similar. To see this, let us define the following index of *expected response time* to capture the overall sluggishness in the response of Δy to Δy^*:

$$\tau \equiv \sum_{k\geq 0} k\mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right], \tag{14}$$

where $\mathrm{E}_t[\cdot]$ denotes expectations conditional on information (that is, values of Δy and Δy^*) known at time t. This index is a weighted sum of the components of the impulse response function, with weights equal to the number of periods that elapse until the corresponding response is observed.[3] 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.

It is easy to show (see Propositions A1 and A2 in the Appendix) that both for the standard Partial Adjustment Model (1) and for the simple lumpy adjustment model (3) we have

$$\tau = \frac{1-\lambda}{\lambda}.$$

More generally, for the model with both gradual and lumpy adjustment described in (9), the expected response to $\Delta\tilde{y}$ satisfies:

$$\tau_{\text{lin}} = \frac{\sum_{k=1}^{K} k\phi_k}{1-\sum_{k=1}^{K}\phi_k},$$

while the expected response to shocks Δy^* is equal to:[4]

$$\tau = \frac{1-\lambda}{\lambda} + \frac{\sum_{k=1}^{K} k\phi_k}{1-\sum_{k=1}^{K}\phi_k}. \tag{15}$$

[3] Note that, for the models at hand, the impulse response is always non-negative and adds up to one. When the impulse response does not add up to one, the definition above needs to be modified to:

$$\tau \equiv \frac{\sum_{k\geq 0} k\mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right]}{\sum_{k\geq 0}\mathrm{E}\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right]}.$$

[4] For the derivation of both expressions for τ see Propositions A1 and A3 in the Appendix.

Let us label:

$$\tau_{\text{lum}} \equiv \frac{1-\lambda}{\lambda}.$$

It follows that that the expected response when both sources of sluggishness are present is the sum of the responses to each one taken separately.

We can now state the implication of Proposition 2 for the estimated expected time of adjustment, $\hat{\tau}$.

Corollary 1 (Fast Adjustment) *Let the assumptions of Proposition 2 hold and let $\hat{\tau}$ denote the (classical) method of moments estimator for τ obtained from OLS estimators of:*

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

Then:

$$\text{plim}_{T\to\infty}\hat{\tau} = \tau_{\text{lin}} \le \tau = \tau_{\text{lin}} + \tau_{\text{lum}}, \tag{16}$$

with a strict inequality for $\lambda < 1$.

Proof See Appendix B.1. ∎

To summarize, the linear approximation for $\Delta\tilde{y}$ (wrongly) suggests no sluggishness whatsoever, so that when this approximation is plugged into the (correct) linear relation between Δy and $\Delta\tilde{y}$, one source of sluggishness is lost. This leads to an expected response time that completely ignores the sluggishness caused by the lumpy component of adjustments.

2.4 An Application: Monetary Policy

Figure 1 depicts the monthly evolution of the intended federal funds rate during the Greenspan era.[5] The infrequent nature of adjustment of this policy variable is evident in the figure. It is also well known that monetary policy interventions often come in gradual steps (see, e.g., Sack (1998) and Woodford (1999)), fitting the description of the model we just characterized.

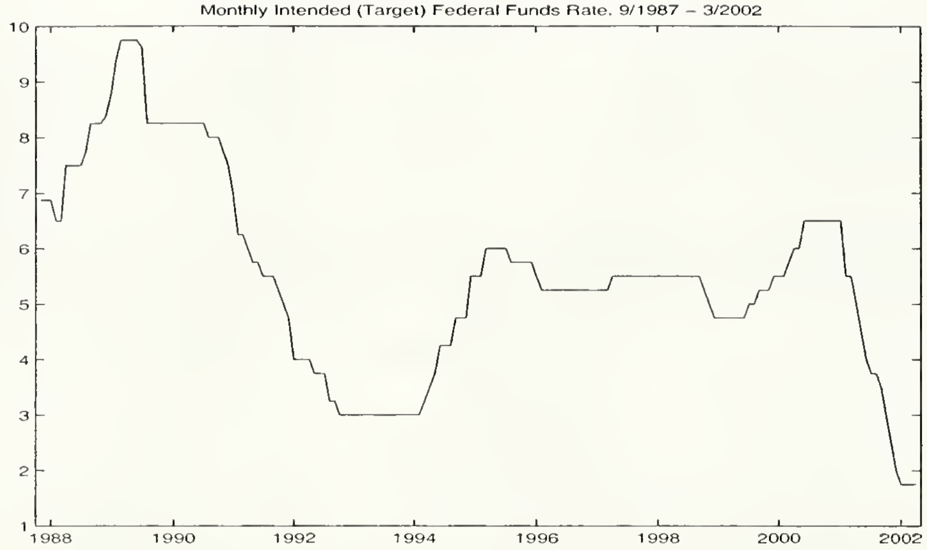
Our goal is to estimate both components of τ, τ_{lin} and τ_{lum}. Regarding the former, we estimate AR processes for Δy with an increasing number of lags, until finding no significant improvement in the goodness-of-fit. This procedure is warranted since $\Delta\tilde{y}$, the omitted regressor, is orthogonal to the lagged Δy's. We obtained an AR(3) process, with τ_{lin} estimated as 2.35 months.

If the lumpy component is relevant, the (absolute) magnitude of adjustments of $\Delta\tilde{y}$ should increase with the number of periods since the last adjustment. The longer the inaction period, the larger the number of shocks in Δy^* to which $\tilde{y}$ has not adjusted, and hence the larger the variance of observed adjustments.

To test this implication of lumpy adjustment, we identified periods with adjustment in $\tilde{y}$ as those where the residual from the linear model takes (absolute) values above a certain threshold, M. Next we partitioned

[5]The findings reported in this section remain valid if we use different sample periods.

Figure 1:



those observations where adjustment took place into two groups. The first group included observations where adjustment also took place in the preceding period, so that the estimated $\Delta\tilde{y}$ only reflects innovations of Δy^* in one period. The second group considered the remaining observations, where adjustments took place after at least one period with no adjustment.

Columns 2 and 3 in Table 2 show the variances of adjustments in the first and second group described above, respectively, for various values of M. Interestingly, the variance when adjustments reflects only one shock is significantly smaller than the variance of adjustments to more than one shock (see column 4). If there were no lumpy component at all ($\lambda = 1$), there would be no systematic difference between both variances, since they would correspond to a random partition of observations where the residual is larger than M. We therefore interpret our findings as evidence in favor of significant lumpy adjustment.[6]

To actually estimate the contribution of the lumpy component to overall sluggishness, we need to estimate the fraction of months where an adjustment in $\tilde{y}$ took place. Since we only observe y, we require some additional information to determine this adjustment rate. If we had a criterion to choose the threshold M, this could be readily done. The fact that the Fed changes rates by multiples of 0.25 suggests that reasonable choices for M are in the neighborhood of this value. Columns 5 and 6 in Table 2 report the values estimated for λ and τ_{lum} for different values of M. For all these cases, the estimated lumpiness is substantial.

An alternative procedure is to extract lumpiness from the behavior of y_t directly. For this approach, we used four criteria: First, if $y_t \neq y_{t-1}$ and $y_{t-1} = y_{t-2}$ then an adjustment of $\tilde{y}$ occurred at t. Similarly if a "reversal" happened at t, that is, if $y_t > y_{t-1}$ and $y_{t-1} < y_{t-2}$ (or $y_t < y_{t-1}$ and $y_{t-1} > y_{t-2}$). By contrast, if $y_t = y_{t-1}$, we assume that no adjustment took place in period t. Finally, if an "acceleration" took place at

[6] Also note that the first order autocorrelation of residuals is -0.002, so that our finding cannot be attributed to this factor either.

Table 2: ESTIMATING THE LUMPY COMPONENT

(1) M	(2) $\text{Var}(\Delta\tilde{y}_t \mid \xi_t = 1, \xi_{t-1} = 1)$	(3) $\text{Var}(\Delta\tilde{y}_t \mid \xi_t = 1, \xi_{t-1} = 0)$	(4) p-value	(5) $\hat{\lambda}$	(6) $\hat{\tau}_{\text{lum}}$
0.150	0.089	0.213	0.032	0.365	1.74
0.175	0.092	0.221	0.018	0.323	2.10
0.200	0.117	0.230	0.051	0.253	2.95
0.225	0.145	0.267	0.060	0.206	3.85
0.250	0.150	0.325	0.034	0.165	5.06
0.275	0.164	0.378	0.016	0.135	6.41
0.300	0.207	0.378	0.045	0.123	7.13
0.325	0.207	0.378	0.045	0.123	7.13
0.350	0.224	0.433	0.041	0.100	9.00

For various values of M (see Column 1), values reported in the remaining columns are as follows. Columns 2 and 3: Estimates of the variance of $\Delta\tilde{y}$, conditional on adjusting, for observations where adjustment also took place the preceding period (column 2) and with no adjustment in the previous period (column 3). Column 4: p-value, obtained via bootstrap, for both variances being the same, against the alternative that the latter is larger. Column 5: Estimates of λ. Column 6: Estimates of τ_{lum}.

t, so that $y_t - y_{t-1} > y_{t-1} - y_{t-2} > 0$ (or $y_t - y_{t-1} < y_{t-1} - y_{t-2} < 0$), we assume that $\tilde{y}$ adjusted at t. With these criteria we can sort 156 out of the 174 months in our sample into (lumpy) adjustment taking place or not. This allows us to bound the (estimated) value of λ between the estimate we obtain by assuming that no adjustment took place in the remaining 18 periods and that in which all of them correspond to adjustments for $\tilde{y}$.

Table 3: MODELS FOR THE INTENDED FEDERAL FUNDS RATE

Time-to-build component				Lumpy component			
ϕ_1	ϕ_2	ϕ_3	$\hat{\tau}_{\text{lin}}$	$\hat{\lambda}_{min}$	$\hat{\lambda}_{max}$	$\hat{\tau}_{\text{lum,min}}$	$\hat{\tau}_{\text{lum,max}}$
0.230	0.080	0.231	2.35	0.221	0.320	2.13	3.53
(0.074)	(0.076)	(0.074)	(0.95)	(0.032)	(0.035)	(0.34)	(0.64)

Reported: Estimation of both components of the sluggishness index τ. Data: Intended (Target) Federal Funds Rate, monthly, 9/1987–3/2002. Standard deviations in parenthesis, obtained via Delta method.

Table 3 summarizes our estimates of the expected response time obtained with this procedure. A researcher who ignores the lumpy nature of adjustments only would consider the AR-component and would infer a value of τ equal to 2.35 months. Yet once we consider infrequent adjustments, the correct estimate of τ is somewhere between 4.48 and 5.88 months. That is, ignoring lumpiness (wrongly) suggests a response to shocks that is approximately twice as fast as the true response.[7]

Consistent with our theoretical results, the bias in the estimated speed of adjustment stems from the

[7]Note that these coincide with estimates obtained for M in the 0.175 to 0.225 range, see Table 2.

infrequent adjustment of monetary policy to news. As shown above, this bias is important, since infrequent adjustment accounts for at least half of the sluggishness in modern U.S. monetary policy.

3 Slow Aggregate Convergence

Could aggregation solve the problem for those variables where lumpiness occurs at the microeconomic level? In the limit, yes. Rotemberg (1987) showed that the aggregate equation resulting from individual actions driven by the Calvo-model indeed converges to the partial-adjustment model. That is, as the number of microeconomic units goes to infinity, estimation of equation (6) for the aggregate does yield the correct estimate of λ, and therefore τ. (Henceforth we return to the simple model without time-to-build).

But not all is good news. In this section, we show that when the speed of adjustment is already slow, the bias vanishes very slowly as the number of units in the aggregate increases. In fact, in the case of investment even aggregating across all U.S. manufacturing establishments is not sufficient to eliminate the bias.

3.1 The Result

Let us define the N-aggregate change at time t, Δy_t^N, as:

$$\Delta y_t^N \equiv \frac{1}{N}\sum_{i=1}^{N} \Delta y_{i,t},$$

where $\Delta y_{i,t}$ denotes the change in the variable of interest by unit i in period t.

Technical Assumptions (Shocks)

Let $\Delta y_{i,t}^* \equiv v_t^A + v_{i,t}^I$, where the absence of a subindex i denotes an element common to all i (i.e, that remains after averaging across all i's). We assume:

1. the v_t^A's are i.i.d. with mean μ_A and variance $\sigma_A^2 > 0$,
2. the $v_{i,t}^I$'s are independent (across units, over time, and with respect to the v^A's), identically distributed with zero mean and variance $\sigma_I^2 > 0$, and
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 $\lambda \in (0,1]$. ∎

As in the single unit case, we now ask whether estimating

$$\Delta y_t^N = (1-\lambda)\Delta y_{t-1}^N + \varepsilon_t, \tag{17}$$

yields a consistent (as T goes to infinity) estimate of λ, when the true microeconomic model is (8). 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 (λ, μ_A, σ_A and σ_I) and, most importantly, N.

Proposition 3 (Aggregate Bias)

Let $\hat{\lambda}^N$ denote the OLS estimator of λ in equation (17). Let T denote the time series length. Then, under the Technical Assumptions,

$$\text{plim}_{T\to\infty}\hat{\lambda}^N = \lambda + \frac{1-\lambda}{1+K}, \tag{18}$$

with

$$K \equiv \frac{\frac{\lambda}{2-\lambda}(N-1)\sigma_A^2 - \mu_A^2}{\sigma_A^2 + \sigma_I^2 + \frac{2-\lambda}{\lambda}\mu_A^2}. \tag{19}$$

It follows that:

$$\lim_{N\to\infty} \text{plim}_{T->\infty}\hat{\lambda}^N = \lambda. \tag{20}$$

Proof See Theorem B1 in the Appendix. ∎

To see the source of the bias and why aggregation reduces it, we begin by writing the first order autocorrelation, ρ_1, as an expression that involves sums and quotients of four different terms:

$$\rho_1 = \frac{\text{Cov}(\Delta y_t^N, \Delta y_{t-1}^N)}{\text{Var}(\Delta y_t^N)} = \frac{[N\text{Cov}(\Delta y_{1,t}, \Delta y_{1,t-1}) + N(N-1)\text{Cov}(\Delta y_{1,t}, \Delta y_{2,t-1})]/N^2}{[N\text{Var}(\Delta y_{1,t}) + N(N-1)\text{Cov}(\Delta y_{1,t}, \Delta y_{2,t})]/N^2}, \tag{21}$$

where the subindex 1 and 2 in Δy denote two different units.

The numerator of (21) includes N (by symmetry identical) first-order autocovariance terms, one for each unit, and $N(N-1)$ (also identical) first-order cross-covariance terms, one for each pair of different units. Likewise, the denominator considers N identical variance terms and $N(N-1)$ identical contemporaneous cross-covariance terms.

From columns 2 and 4 in Table 4 we observe that the cross-covariance terms under PAM and lumpy adjustment are the same.[8] Since these terms will dominate for sufficiently large N —there are $N(N-1)$ of them, compared to N additional terms— it follows that the bias vanishes as N goes to infinity.

[8] This is somewhat remarkable, since the underlying processes are quite different. For example, consider the first-order cross-covariance term. In the case of PAM, adjustments at all lags contribute to the cross-covariance term:

$$\begin{aligned}
\text{Cov}(\Delta y_{1,t}, \Delta y_{2,t-1}) &= \text{Cov}(\lambda \sum_{k\geq 0}(1-\lambda)^k \Delta y^*_{1,t-k}, \lambda \sum_{l\geq 0}(1-\lambda)^l \Delta y^*_{2,t-1-l}) \\
&= \sum_{k=l+1} \lambda^2 (1-\lambda)^{k+l+1} \text{Cov}(\Delta y^*_{1,t-k}, \Delta y^*_{2,t-1-l}) \\
&= \left[\sum_{l\geq 0} \lambda^2 (1-\lambda)^{2l+1}\right] \sigma_A^2 = \frac{\lambda}{2-\lambda}(1-\lambda)\sigma_A^2.
\end{aligned}$$

By contrast, in the case of the lumpy adjustment model, the non-zero terms obtained when calculating the covariance between $\Delta y_{1,t}$ and $\Delta y_{2,t-1}$ are due to aggregate shocks included both in the adjustment of unit 1 (in t) and unit 2 (in $t-1$). Idiosyncratic shocks

Table 4: CONSTRUCTING THE FIRST ORDER CORRELATION

		(1) $\mathrm{Cov}(\Delta y_{1,t},\Delta y_{1,t-1})$	(2) $\mathrm{Cov}(\Delta y_{1,t},\Delta y_{2,t-1})$	(3) $\mathrm{Var}(\Delta y_{1,t})$	(4) $\mathrm{Cov}(\Delta y_{1,t},\Delta y_{2,t})$
(1)	PAM:	$\frac{\lambda}{2-\lambda}(1-\lambda)(\sigma_A^2+\sigma_I^2)$	$\frac{\lambda}{2-\lambda}(1-\lambda)\sigma_A^2$	$\frac{\lambda}{2-\lambda}(\sigma_A^2+\sigma_I^2)$	$\frac{\lambda}{2-\lambda}\sigma_A^2$
(2)	Lumpy ($\mu_A=0$):	0	$\frac{\lambda}{2-\lambda}(1-\lambda)\sigma_A^2$	$\sigma_A^2+\sigma_I^2$	$\frac{\lambda}{2-\lambda}\sigma_A^2$
(3)	Lumpy ($\mu_A\neq 0$):	$-(1-\lambda)\mu_A^2$	$\frac{\lambda}{2-\lambda}(1-\lambda)\sigma_A^2$	$\sigma_A^2+\sigma_I^2+\frac{2(1-\lambda)}{\lambda}\mu_A^2$	$\frac{\lambda}{2-\lambda}\sigma_A^2$

However, the underlying bias may remain significant for relatively large values of N. From Table 4 it follows that the bias for the estimated first-order autocorrelation originates from the autocovariance terms included in both the numerator and denominator. The first-order autocovariance term in the numerator is zero for the lumpy adjustment model, while it is positive under PAM (this is the bias we discussed in Section 2). And even though the number of terms with this bias is only N, compared with $N(N-1)$ cross-covariance terms with no bias, the missing terms are proportional to $\sigma_A^2+\sigma_I^2$, while those that are included are proportional to σ_A^2, which is considerably smaller in all applications. This suggests that the bias remains significant for relatively large values of N (more on this below) and that this bias rises with the relative importance of idiosyncratic shocks.

There is a second source of bias once $N>1$, related to the variance term $\mathrm{Var}(\Delta y_{1,t})$ in the denominator of the first-order correlation in (21). While under PAM this variance is increasing in λ, varying between 0 (when $\lambda=0$) and $\sigma_A^2+\sigma_I^2$ (when $\lambda=1$), when adjustment is lumpy this variance attains the largest possible value under PAM, $\sigma_A^2+\sigma_I^2$, independent of the underlying adjustment speed λ. This suggests that the bias is more important when adjustment is fairly infrequent.[9]

Substituting the terms in the numerator and denominator of (21) by the expressions in the second row of

are irrelevant as far as the covariance is concerned. It follows that:

$$\mathrm{Cov}[\Delta y_{1,t},\Delta y_{2,t-1}|\xi_{1,t}=1,\xi_{2,t-1}=1,l_{1,t},l_{2,t-1}]=\min(l_{1,t}-1,l_{2,t-1})\sigma_A^2,$$

and averaging over $l_{1,t}$ and $l_{2,t-1}$, both of which follow (independent) Poisson processes, we obtain:

$$\mathrm{Cov}(\Delta y_{1,t},\Delta y_{2,t-1})=\frac{\lambda}{2-\lambda}(1-\lambda)\sigma_A^2, \tag{22}$$

which is the expression obtained under PAM.

[9]To further understand why $\mathrm{Var}(\Delta y_{1,t})$ can be so much larger in a Calvo model than with partial adjustment, we compare the contribution to this variance of shocks that took place k periods ago, $V_{k,t}$.

For PAM we have

$$\mathrm{Var}(\Delta y_{1,t})=\mathrm{Var}(\sum_{k\geq 0}\lambda(1-\lambda)^k\Delta y^*_{1,t-k})=\lambda^2\sum_{k\geq 0}(1-\lambda)^{2k}\mathrm{Var}(\Delta y^*_{1,t-k}), \tag{23}$$

Table 4, and dividing numerator and denominator by $N(N-1)\lambda/(2-\lambda)$ leads to:

$$\rho_1 = \frac{1-\lambda}{1+\frac{2-\lambda}{\lambda(N-1)}\left(1+\frac{\sigma_I^2}{\sigma_A^2}\right)}. \tag{24}$$

This expression confirms our discussion. It illustrates clearly that the bias is increasing in σ_I/σ_A and decreasing in λ and N.

Finally, we note that a value of $\mu_A \neq 0$ biases the estimates of the speed of adjustment even further. The reason for this is that it introduces a sort of "spurious" negative correlation in the time series of $\Delta y_{1,t}$. Whenever the unit does not adjust, its change is, in absolute value, below the mean change. When adjustment finally takes place, pent-up adjustments are undone and the absolute change, on average, exceeds $|\mu_A|$. The product of these two terms is clearly negative, inducing negative serial correlation.[10]

Summing up, the bias obtained when estimating λ with standard partial adjustment regressions can be expected to be significant when either σ_A/σ_I, N or λ is small, or $|\mu_A|$ is large. Figure 2 illustrates how $\hat{\lambda}^N$ converges to λ. The baseline parameters (solid line) are $\mu_A = 0$, $\lambda = 0.20$, $\sigma_A = 0.03$ and $\sigma_I = 0.24$. The solid line depicts the percentage bias as N grows. For $N = 1{,}000$, the bias is above 100%; by the time $N = 10{,}000$, it is slightly above 20%. The dash-dot line increases σ_A to 0.04, speeding up convergence. By contrast, the dash-dot line considers $\mu_A = 0.10$, which slows down convergence. Finally, the dotted line shows the case where λ doubles to 0.40, which also speeds up convergence.

Corollary 2 (Slow Convergence)

The bias in the estimator of the adjustment speed is increasing in σ_I and $|\mu_A|$ and decreasing in σ_A, N and λ. Furthermore, the four parameters mentioned above determine the bias of the estimator via a decreasing expression of K.[11]

Proof Trivial. ∎

---

so that

$$V_{k,t}^{\text{PAM}} = \lambda^2(1-\lambda)^{2k}(\sigma_A^2+\sigma_I^2).$$

By contrast, in the case of lumpy adjustment we have:

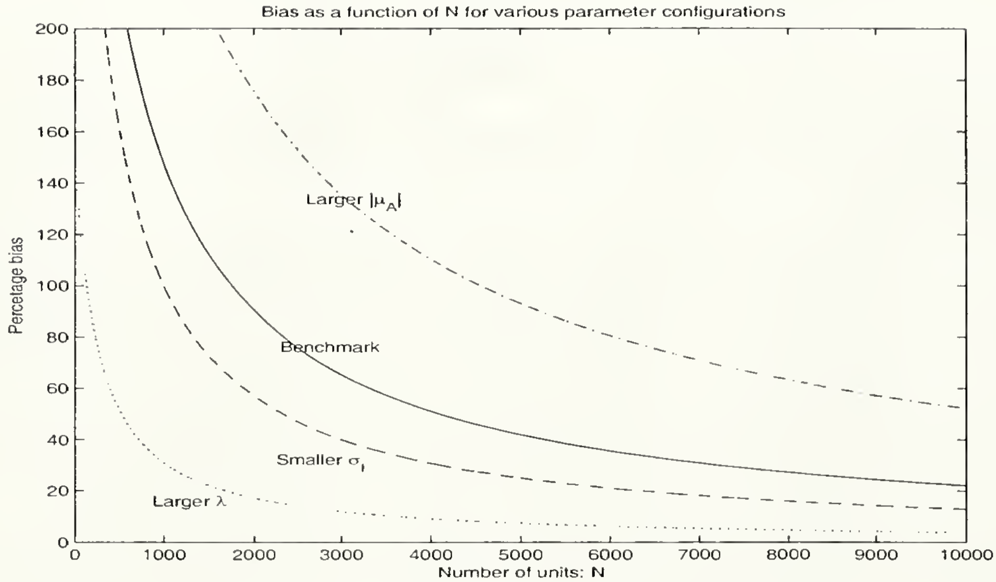
$$\begin{aligned} V_k^{\text{lumpy}} &= \Pr\{l_{1,t}=k\}\mathrm{Var}(\Delta y_{1,t}|l_{1,t}=k) \\ &= \Pr\{l_{1,t}=k\}[\lambda\mathrm{Var}(\Delta y_{1,t}|l_{1,t}=1,\xi_{1,t}=1)+(1-\lambda)\mathrm{Var}(\Delta y_{1,t}|l_{1,t}=0,\xi_t=0)] \\ &= \lambda(1-\lambda)^{k-1}[\lambda k(\sigma_A^2+\sigma_I^2)]. \end{aligned}$$

$V_{k,t}$ is much larger under the lumpy adjustment model than under PAM. With infrequent adjustment the relevant conditional distribution is a mixture of a mass at zero (corresponding to no adjustment at all) and a normal distribution with a variance that grows linearly with k (corresponding to adjustment in t). Under PAM, by contrast, $V_{k,t}$ is generated from a a normal distribution with zero mean and variance that decreases with k.

[10] We also have that $\mu_A \neq 0$ further increases the bias due to the variance term, see entry (3,3) in Table 4

[11] The results for σ_I and λ may not hold if $|\mu_A|$ is large. For the results to hold we need $N > 1 + (2-\lambda)\mu_A^2/\lambda\sigma_A^2$. When $\mu_A = 0$ this is equivalent to $N \geq 1$ and therefore is not binding.

Figure 2:



3.2 Applications

Figure 2 shows that the bias in the estimate of the speed of adjustment is likely to remain significant, even when estimated with very aggregated data. In this section we provide concrete examples based on estimates for U.S. employment, investment, and price dynamics. These series are interesting because there is extensive evidence of their infrequent adjustment at the microeconomic level.

Let us start with U.S. manufacturing employment. We use the parameters estimated by Caballero, Engel, and Haltiwanger (1997) with quarterly Longitudional Research Datafile (LRD) data. Table 5 shows that even when $N = 1,000$, which corresponds to many more establishments than in a typical two-digit sector of the LRD, the bias remains above 40 percent. That is, estimated speeds of adjustment at the sectoral level are likely to be significantly faster than the true speed of adjustment.

The good news in this case is that for $N = 10,000$, which is about the size of the continuous sample in the LRD, the bias essentially vanishes.

The results for prices, reported in Table 6, are based on the estimate of λ, μ_A and σ_A from Bils and Klenow (2002), while σ_I is consistent with that found in Caballero et al (1997).[12] The table shows that the

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

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

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

Table 5: SLOW CONVERGENCE: EMPLOYMENT
Average $\hat{\lambda}$

		Number of agents (N)			
		100	1,000	10,000	∞
Number of Time Periods (T)	35	0.901	0.631	0.523	
	200	0.852	0.548	0.436	
	∞	0.844	0.532	0.417	0.400

Reported: average of OLS estimates of λ, obtained via simulations. Number of simulations chosen to ensure that numbers reported have a standard deviation less than 0.002. Case $T = \infty$ calculated from Proposition 3. Simulation parameters: λ: 0.40, $\mu_A = 0.005$, $\sigma_A = 0.03$, $\sigma_I = 0.25$. Quarterly data, from Caballero et al. (1997).

bias remains significant even for $N = 10,000$. In this case, the main reason for the stubborn bias is the high value of σ_I/σ_A.

Table 6: SLOW CONVERGENCE: PRICES
Average $\hat{\lambda}$

		Number of agents (N)			
		100	1,000	10,000	∞
Number of Time Periods (T)	60	0.935	0.614	0.351	
	500	0.908	0.542	0.279	
	∞	0.902	0.533	0.269	0.220

Reported: average OLS estimates of λ, obtained via simulations. Number of simulations chosen to ensure that numbers reported have a standard deviation less than 0.002. Case $T = \infty$ calculated from Proposition 3. Simulation parameters: λ: 0.22 (monthly data, Bils and Klenow, 2002), $\mu_A = 0.003$, $\sigma_A = 0.0054$, $\sigma_I = 0.048$.

Finally, Table 7 reports the estimates for equipment investment, the most sluggish of the three series. The estimate of λ, μ_A and σ_A, are from Caballero, Engel, and Haltiwanger (1995), and σ_I is consistent with that found in Caballero et al (1997).[13] Here the bias remains very large and significant throughout. Even when $N = 10,000$, the estimated speed of adjustment exceeds the actual speed by more than 80 percent. The

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

reasons for this is the combination of a low λ, a high μ_A (mostly due to depreciation), and a large σ_I (relative to σ_A).

Table 7: SLOW CONVERGENCE: INVESTMENT
Average $\hat{\lambda}$

		Number of agents (N)			
		100	1,000	10,000	∞
Number of Time Periods (T)	15	1.060	0.950	0.665	
	50	1.004	0.790	0.400	
	200	0.985	0.723	0.305	
	∞	0.979	0.696	0.274	0.150

Reported: average of OLS estimates of λ, obtained via simulations. Number of simulations chosen to ensure that numbers reported have a standard deviation less than 0.002. Simulation parameters: λ: 0.15 (annual data, from Caballero et al, 1995), $\mu_A = 0.12$, $\sigma_A = 0.056$, $\sigma_I = 0.50$.

We have assumed throughout that Δy^* is i.i.d. Aside from making the results cleaner, it should be apparent from the time-to-build extension in Section 2 that adding further serial correlation does not change the essence of our results. In such a case, the cross correlations between contiguous adjustments are no longer zero, but the bias we have described remains. In any event, for each of the applications in this subsection, there is evidence that the i.i.d. assumption is not farfetched (see, e.g., Caballero et al [1995, 1997], Bils and Klenow [2002]).

4 Fragile Solutions: Biased Regressions and ARMA Correction

Can we fix the problem while remaining within the class of linear time-series models? In this section, we show that in principle this is possible, but in practice it is unlikely (especially for small N).

4.1 Biased Regressions

So far we have assumed that the speed of adjustment 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 (2), the estimating equation is:

$$\Delta y_t = (1-\lambda)\Delta y_{t-1} + \lambda \Delta y_t^* + e_t, \tag{25}$$

with some proxy available for the regressor Δy^*.

Equation (25) hints at a procedure for solving the problem. Since the regressors are orthogonal, λ in principle can be estimated directly from the parameter estimate associated with Δy_t^*, while dropping the constraint that the sum of the coefficients on the right hand side add up to one. Of course, if the econometrician does impose the latter constraint, then the estimate of λ will be some weighted average of an unbiased and a biased coefficient, and hence will be biased as well. We summarize these results in the following proposition.

Proposition.4 (Bias with Regressors)

With the same notation and assumptions as in Proposition 3, consider the following equation:

$$\Delta y_t^N = b_0 \Delta y_{t-1}^N + b_1 \Delta y_t^* + e_t, \tag{26}$$

where Δy_t^* denotes the average shock in period t, $\sum \Delta y_{i,t}^*/N$. Then, if (26) is estimated via OLS, and K defined in (19),

(i) without any restrictions on b_0 and b_1:

$$\begin{aligned} plim_{T\to\infty}\hat{b}_0 &= \frac{K}{1+K}(1-\lambda), && (27) \\ plim_{T\to\infty}\hat{b}_1 &= \lambda, && (28) \end{aligned}$$

(ii) imposing $b_0 = 1 - b_1$:[14]

$$plim_{T\to\infty}\hat{b}_1 = \lambda + \frac{\lambda(1-\lambda)}{2(K+\lambda)}.$$

In particular, for $N = 1$ and $\mu_A = 0$:

$$plim_{T\to\infty}\hat{b}_1 = \lambda + \frac{1-\lambda}{2}.$$

Proof See Corollary B1 in the Appendix. ∎

Of course, in practice the "solution" above is not very useful. First, the econometrician seldom observes Δy^* exactly, and (at least) the scaling parameters need to be estimated. In this situation, the coefficient estimate on the contemporaneous proxy for Δy^* is no longer useful for estimating λ, and the latter must be estimated from the serial correlation of the regression, bringing back the bias. Second, when the econometrician does observe Δy^*, the adding up constraint typically is linked to homogeneity and long-run conditions

[14]The expression that follows is a weighted average of the unbiased estimator λ and the biased estimator in the regression without Δy^* as a regressor (Proposition 3). The weight on the biased estimator is $\lambda K/2(K+\lambda)$, which corresponds to the harmonic mean of λ and K.

that a researcher often will be reluctant to drop (see below).

Fast Micro – Slow Macro?: A Price-Wage Equation Application

In an intriguing article, Blanchard (1987) reached the conclusion that the speed of adjustment of prices to cost changes is much faster at the disaggregate than the aggregate level. More specifically, he found that prices adjust faster to wages (and input prices) at the two-digit level than at the aggregate level. His study considered seven manufacturing sectors and estimated equations analogous to (26), with sectoral prices in the role of y, and both sector-specific wages and input prices as regressors (the y^*). The classic homogeneity condition in this case, which was imposed in Blanchard's study, is equivalent in our setting to $b_0 + b_1 = 1$.

Blanchard's preferred explanation for his finding was based on the slow transmission of price changes through the input-output chain. This is an appealing interpretation and likely to explain some of the difference in speed of adjustment at different levels of aggregation. However, one wonders how much of the finding can be explained by biases like those described in this paper. We do not attempt a formal decomposition but simply highlight the potential size of the bias in price-wage equations for realistic parameters.

Matching Blanchard's framework to our setup, we know that his estimated sectoral λ is approximately 0.18 while at the aggregate level it is 0.135.[15]

Table 8: Biased Speed of Adjustment: Price-Wage Equations
Average $\hat{\lambda}$

		Number of firms (N)					
		100	500	1,000	5,000	10,000	∞
No. of Time Periods, T:	250	0.416	0.235	0.194	0.148	0.142	0.135
	∞	0.405	0.239	0.194	0.148	0.142	0.135

Reported: For $T = 250$, average estimate of λ obtained via simulations. Number of simulations chosen to ensure that estimates reported have a standard deviation less than 0.004. For $T = \infty$: calculated from (28). Simulation parameters: λ: 0.135 and $T = 250$ (from Blanchard, 1987, monthly data), $\mu_A = 0.003$, $\sigma_A = 0.0054$ and $\sigma_I = 0.048$ as in Table 6.

Table 8 reports the bias obtained when estimating the adjustment speed from sectoral price-wage equations. It assumes that the true speed of adjustment, λ, is 0.135, and considers various values for the number of firms in the sector. The table shows that for reasonable values of N there is a significant upward bias in the estimated value of λ, certainly enough to include Blanchard's estimates.[16]

[15] We obtained these estimates by matching the cumulative impulse responses reported in the first two columns of Table 8 in Blanchard (1987) for 5, 6 and 7 lags. For the sectoral speeds we obtain, respectively, 0.167, 0.182 and 0.189, while for the aggregate speed we obtain 0.137, 0.121 and 0.146. The numbers in the main text are the average λ's obtained this way.

[16] The estimated speed of adjustment is close to 0.18 for $N = 1,000$.

4.2 ARMA Corrections

Let us go back to the case of unobserved Δy^*. Can we fix the bias while remaining within the class of linear ARMA models? In the first part of this subsection, we show that this is indeed possible. Essentially, the correction amounts to adding a nuisance MA term that "absorbs" the bias.

However, the second part of this subsection warns that this nuisance parameter needs to be ignored when estimating the speed of adjustment. This is not encouraging, because in practice the researcher is unlikely to know when he should or should not drop some of the MA terms before simulating (or drawing inferences from) the estimated dynamic model.

On the constructive side, nonetheless, we show that when N is sufficiently large, even if we do not ignore the nuisance MA parameter, we obtain better —although still biased— estimates of the speed of adjustment than with the simple partial adjustment model.

4.2.1 Nuisance Parameters and Bias Correction

Let us start with the positive result.

Proposition 5 (Bias Correction) *Let the Technical Assumptions (see page 10) hold.*[17] *Then Δy_t^N follows an ARMA(1,1) process with autoregressive parameter equal to $1-\lambda$. Thus, adding an MA(1) term to the standard partial adjustment equation (2):*

$$\Delta y_t^N = (1-\lambda)\Delta y_{t-1}^N + v_t - \theta v_{t-1}, \tag{29}$$

and denoting by $\hat{\lambda}^N$ any consistent estimator of one minus the AR-coefficient in the equation above, we have that:

$$plim_{T\to\infty}\hat{\lambda}^N = \lambda.$$

The moving average coefficient, θ, is a "nuisance" parameter that depends on N (it converges to zero as N tends to infinity), μ_A, σ_A and σ_I. We have that:

$$\theta = \frac{1}{2}(L - \sqrt{L^2-4}) > \theta,$$

with

$$L \equiv \frac{2+\lambda(2-\lambda)(K-1)}{1-\lambda},$$

and K defined in (19).

Proof See Theorem B1 in the Appendix. ∎

[17] Strictly speaking, to avoid the case where the AR and MA coefficients coincide, we need to rule out the knife-edge case $N-1 = (2-\lambda)\mu_A^2/\lambda\sigma_A^2$. In particular, when $\mu_A = 0$ this amounts to assuming $N > 1$.

The proposition shows that adding an MA(1) term to the standard partial adjustment equation eliminates the bias. This rather surprising result is valid for any level of aggregation. However, in practice this correction is not robust for small N, as the MA and AR coefficients are very similar in this case (coincidental reduction).[18] Also, as with all ARMA estimation procedures, the time series needs to be sufficiently long (typically $T > 100$) to avoid a significant small sample bias.

Table 9: ADJUSTMENT SPEED τ: WITH AND WITHOUT MA CORRECTION

		Number of agents (N)		
		100	1,000	10,000
	AR(1):	0.18	0.88	1.40
Employment	ARMA(1,1):	0.65	1.29	1.47
	ARMA(1,1), ignoring MA:	1.50	1.50	1.50
	AR(1):	0.11	0.88	2.72
Prices	ARMA(1,1):	1.27	2.83	3.43
	ARMA(1,1), ignoring MA:	3.55	3.55	3.55
	AR(1):	0.02	0.44	2.65
Investment	ARMA(1,1):	0.77	3.92	5.29
	ARMA(1,1), ignoring MA:	5.67	5.67	5.67

Reported: theoretical value of τ, ignoring small sample bias ($T = \infty$). "AR(1)" and "ARMA(1,1)" refer to values of τ obtained from AR(1) and ARMA(1,1) representations. "ARMA(1,1) ignoring MA" refers to estimate obtained using ARMA(1,1) representation, but ignoring the MA term. The results in Propositions 3 and 5 were used to calculate the expressions for τ. Parameter values are those reported in Tables 4, 5 and 6.

Next we illustrate the extent to which our ARMA correction estimates the correct response time in the applications to employment, prices, and investment considered in Section 3. We begin by noting that the expected response time inferred without dropping the MA term is:[19]

$$\tau_{\mathrm{ma}} = \frac{1-\lambda}{\lambda} - \frac{\theta}{1-\theta} < \frac{1-\lambda}{\lambda} = \tau,$$

where $\theta > 0$ was defined in Proposition 5, τ denotes the correct expected response and τ_{ma} the expected response that is inferred from a non-parsimonious ARMA process (it could be an MA(∞), an AR(∞) or, in our particular case, an ARMA(1,1)).

The third row in each of the applications in Table 9 illustrates the main result. The estimate of τ, when the nuisance term is used in estimation but dropped for τ-calculations, is unbiased in all the cases, regardless of the value of N (note that in order to isolate the biases that concern us we have assumed $T = \infty$).

[18] For example, if $\mu_A = 0$, we have that the AR and MA term are identical for $N = 1$.

[19] See Proposition A1 in the Appendix for a derivation.

The first and second rows (in each application) show the biased estimates. The former repeats our basic result while the latter illustrates the problems generated by not dropping the nuisance MA term (τ_{ma}). The bias from not dropping the MA term is smaller than that from inferring τ from the first order autocorrelation,[20] yet it remains significant even at fairly high levels of aggregation (e.g., $N = 1{,}000$).

5 Conclusion

The practice of approximating dynamic models with linear ones is widespread and useful. However, it can lead to significant overestimates of the speed of adjustment of sluggish variables. The problem is most severe when dealing with data at low levels of aggregation or single-policy variables. For once, macroeconomic data seem to be better than microeconomic data.

Yet this paper also shows that the disappearance of the bias with aggregation can be extremely slow. For example, in the case of investment, the bias remains above 80 percent even after aggregating across all continuous establishments in the LRD ($N = 10{,}000$).

While the researcher may think that at the aggregate level it does not matter much which microeconomic adjustment-cost model generates the data, it does matter greatly for (linear) estimation of the speed of adjustment.

What happened to Wold's representation, according to which any stationary, purely non-deterministic, process admits an (eventually infinite) MA representation? Why, as illustrated by the analysis at the end of Section 4, do we obtain an upward biased speed of adjustment when using this representation for the stochastic process at hand? The problem is that Wold's representation expresses the variable of interest as a distributed lag (and therefore linear function) of innovations that are the one-step-ahead *linear* forecast errors. When the relation between the macroeconomic variable of interest and shocks is non-linear, as is the case when adjustment is lumpy, Wold's representation misidentifies the underlying shock, leading to biased estimates of the speed of adjustment.

Put somewhat differently, when adjustment is lumpy, Wold's representation identifies the correct expected response time to the wrong shock. Also, and for the same reason, the impulse response more generally will be biased. So will many of the dynamic systems estimated in VAR style models, and the structural tests that derive from such systems. We are currently working on these issues.

[20] This can be proved formally based on the expressions derived in Theorem B1 and Proposition A1 in the appendix.

References

[1] Bils, Mark and Peter J. Klenow, "Some Evidence on the Importance of Sticky Prices," NBER WP # 9069, July 2002.

[2] Blanchard, Olivier J., "Aggregate and Individual Price Adjustment," *Brookings Papers on Economic Activity 1,* 1987, 57–122.

[3] Box, George E.P., and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day, 1976.

[4] 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.

[5] 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.

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

[7] 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.

[8] Goodfriend, Marvin, "Interest-rate Smoothing and Price Level Trend Stationarity," *Journal of Monetary Economics, 19*, 19987, pp. 335-348.

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

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

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

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

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

Appendix

A The Expected Response Time Index: τ

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

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

with $\psi_0 = 1$, $\sum_{k\geq 0}\psi_k^2 < \infty$, *the* 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.*

Define:

$$I_k \equiv \mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \varepsilon_t}\right] \qquad \text{and} \qquad \tau \equiv \frac{\sum_{k\geq 0} k I_k}{\sum_{k\geq 0} I_k}. \tag{30}$$

Then:

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

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

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

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

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

Define τ *as in (30). 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:

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

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

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

Proposition A2 (τ for a Lumpy Adjustment Process) *Consider* Δy_t *in the simple lumpy adjustment model (8) and* τ *defined in (14). Then* $\tau = (1-\lambda)/\lambda$.[21]

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

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

$$I_k \equiv \mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right] = \Pr\{\xi_{t+k}=1 \,,\, \xi_{t+k-1}=\xi_{t+k-2}=\ldots=\xi_t=0\} = \lambda(1-\lambda)^k. \tag{31}$$

The expression for τ now follows easily. ∎

Proposition A3 (τ for a Process With Time-to-build and Lumpy Adjustments) *Consider the process Δy_t with both gradual and lumpy adjustments:*

$$\Delta y_t = \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + (1 - \sum_{k=1}^{K} \phi_k) \Delta \tilde{y}_t, \tag{32}$$

with

$$\Delta \tilde{y}_t = \xi_t \sum_{k=0}^{l_t - 1} \Delta y_{t-k}^*, \tag{33}$$

where Δy^* is i.i.d. with zero mean and variance σ^2.

Define τ by:

$$\tau \equiv \frac{\sum_{k\geq 0} k \mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right]}{\sum_{k\geq 0} \mathrm{E}\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right]}.$$

Then:

$$\tau = \frac{\sum_{k=1}^{K} k\phi_k}{1-\sum_{k=1}^{K}\phi_k} + \frac{1-\lambda}{\lambda}.$$

Proof Note that:

$$I_k \equiv \mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta y_t^*}\right] = \sum_{j=0}^{k} \mathrm{E}_t\left[\frac{\partial \Delta y_{t+k}}{\partial \Delta \tilde{y}_{t+j}} \frac{\partial \Delta \tilde{y}_{t+j}}{\partial \Delta y_t^*}\right] = \sum_{j=0}^{k} \mathrm{E}_t\left[G_{k-j} \frac{\Delta \tilde{y}_{t+j}}{\Delta y_t^*}\right] = \sum_{j=0}^{k} G_{k-j} H_j, \tag{34}$$

where, from Proposition A1 and (31) we have that the G_k are such that $G(z) \equiv \sum_{k\geq 0} G_k z^k = 1/\Phi(z)$, and $H_k = \lambda(1-\lambda)^k$. Define $H(z) \equiv \sum_{k\geq 0} H_k z^k$ and $I(z) = G(z)H(z)$. Noting that the coefficient of z^k in the infinite series $I(z)$ is equal to I_k in (34), we have:

$$\tau = \frac{I'(1)}{I(1)} = \frac{G'(1)}{G(1)} + \frac{H'(1)}{H(1)} = \frac{\sum_{k=1}^{K} k\phi_k}{1-\sum_{k=1}^{K}\phi_k} + \frac{1-\lambda}{\lambda}. \quad ∎$$

B Bias Results

B.1 Results in Section 2

In this subsection we prove Proposition 2 and Corollary 1. Proposition 1 is a particular case of Proposition 3, which is proved in Section B.2. The notation and assumptions are the same as in Proposition A3.

Proof of Proposition 2 and Corollary 1 The equation we estimate is:

$$\Delta y_t = \sum_{k=1}^{K+1} a_k \Delta y_{t-k} + v_t, \tag{35}$$

while the true relation is that described in (32) and (33).

An argument analogous to that given in Section 2.2 shows that the second term on the right hand side of (32), denoted by w_t in what follows, is uncorrelated with Δy_{t-k}, $k \geq 1$. It follows that estimating (35) is equivalent to estimating (32) 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:

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

The expression for $\text{plim}_{T\to\infty}\hat{\tau}$ now follows from Proposition A3. ∎

B.2 Results in Sections 3 and 4

In this section we prove Propositions 3, 4, and 5. The notation and assumptions are those in Proposition 3. The proof proceeds via a series of lemmas. Propositions 3 and 5 are proved in Theorem B1, while Proposition 4 is proved in Corollary B1.

Lemma B1 *Assume X_1 and X_2 are i.i.d. geometric random variables with parameter λ, so that $\Pr\{X = k\} = \lambda(1-\lambda)^{k-1}, k = 1,2,3,\ldots$ Then:*

$$\begin{aligned} \text{E}[X_i] &= \frac{1}{\lambda}, \\ \text{Var}[X_i] &= \frac{1-\lambda}{\lambda^2}. \end{aligned}$$

In particular, the $l_{i,t}$'s (defined in the main text) are all geometric random variables with parameter λ. Furthermore, $l_{i,t}$ and $l_{j,s}$ are independent if $i \neq j$.

Next define, for any integer s larger or equal than zero:

$$M_s = \begin{cases} 0 & \text{if } X_1 \leq s, \\ \min(X_1 - s, X_2) & \text{if } X_1 > s. \end{cases}$$

Then

$$\mathrm{E}[M_s] = \frac{(1-\lambda)^s}{\lambda(2-\lambda)}.$$

Proof The expressions for $\mathrm{E}[X_i]$ and $\mathrm{Var}[X_i]$ are well known. The properties of the $l_{i,t}$'s are also trivial. To derive the expression for $\mathrm{E}[M_s]$, denote by $F(k)$ and $f(k)$ the cumulative distribution and probability functions common to X_1 and X_2. Then:

$$\begin{aligned}
\Pr\{M_s = k\} &= \Pr\{X_1 - s = k, X_2 \geq k+1\} + \Pr\{X_2 = k, X_1 - s \geq k+1\} + \Pr\{X_1 - s = k, X_2 = k\} \\
&= f(s+k)[1-F(k)] + f(k)[1-F(k+s)] + f(k)f(k+s),
\end{aligned}$$

and, since $1 - F(k) = (1-\lambda)^k$, with some algebra we obtain:

$$\Pr\{M_s = k\} = \lambda(2-\lambda)(1-\lambda)^{s+2k-2}.$$

Using this expression to calculate $\mathrm{E}[M_s]$ via $\sum_{k\geq 1} k \Pr\{M_s = k\}$ completes the proof. ∎

Lemma B2 *For any strictly positive integer s,*

$$\mathrm{Cov}(\Delta y_{i,t}, \Delta y_{i,t+s}) = -(1-\lambda)^s \mu_A^2.$$

Proof We have:

$$\begin{aligned}
\mathrm{Cov}(\Delta y_{t+s}, \Delta y_t) &= \mathrm{E}[\Delta y_{t+s}\Delta y_t] - \mu_A^2 \\
&= \sum_{k=1}^{\infty}\sum_{q=1}^{s} \mathrm{E}[\Delta y_{t+s}\Delta y_t | l_{t+s} = q, l_t = k, \xi_{t+s} = 1, \xi_t = 1] \Pr\{l_{t+s} = q, l_t = k, \xi_{t+s} = 1, \xi_t = 1\} - \mu_A^2 \\
&= \mu_A^2 \sum_{k=1}^{\infty}\sum_{q=1}^{s} qk \Pr\{l_{t+s} = q, l_t = k, \xi_{t+s} = 1, \xi_t = 1\} - \mu_A^2,
\end{aligned}$$

where in the second (and only non-trivial) step we add up over a partition of the set of outcomes where $\Delta y_{t+s}\Delta y_t \neq 0$.

The expression above, combined with:

$$\Pr\{l_{t+s} = q, l_t = k, \xi_{t+s} = 1, \xi_t = 1\} = \begin{cases} \lambda^4(1-\lambda)^{k+q-2}, & \text{if } q = 1, \ldots, s-1, \\ \lambda^3(1-\lambda)^{k+q-2}, & \text{if } q = s, \end{cases}$$

and some patient algebra completes the proof. ∎

Lemma B3 *For $q \neq r$ and any integer s larger or equal than zero we have:*

$$\mathrm{Cov}(\Delta y_{q,t+s}, \Delta y_{r,t}) = \frac{\lambda}{2-\lambda}(1-\lambda)^s \sigma_A^2.$$

Proof Denote $v_{i,t} \equiv \Delta y_{i,t}^* = v_t^A + v_{i,t}^I$. Then:

$$\mathrm{E}[\Delta y_{q,t+s}\Delta y_{r,t} | l_{q,t+s}, l_{r,t}] = \mathrm{E}[\xi_{q,t+s}(\sum_{j=0}^{l_{q,t+s}-1} v_{q,t+s-j})\xi_{r,t}(\sum_{k=0}^{l_{r,t}-1} v_{r,t-k}) | l_{q,t+s}, l_{r,t}]$$

$$
\begin{aligned}
&= \lambda^2 \sum_{j=0}^{l_{q,t+s}-1} \sum_{k=0}^{l_{r,t}-1} \mathrm{E}[\nu^A_{t+s-j}\nu^A_{t-k}] \\
&= \lambda^2 l_{q,t+s} l_{r,t} \mu_A^2 \; + \; \lambda^2 M_s(l_{q,t+s}, l_{r,t}) \sigma_A^2.
\end{aligned}
$$

Where the first identity follows from the definition of the $\Delta y_{i,t}$'s, the second from conditioning on the four possible combinations of values of $\xi_{q,t+s}$ and $\xi_{r,t}$, and $M_s(l_{q,t+s}, l_{r,t})$ denotes the random variable analogous to M_s in Lemma B1, based on the i.i.d. geometric random variables $l_{q,t+s}$ and $l_{r,t}$. The remainder of the proof is based on the expressions in Lemma B1 and straightforward algebra. ∎

Lemma B4 *We have:*

$$\mathrm{Var}(\Delta y_{i,t}) = \frac{2(1-\lambda)}{\lambda}\mu_A^2 + \sigma_A^2 + \sigma_I^2.$$

Proof The proof is based on calculating both terms on the right hand side of the well known identity:

$$\mathrm{Var}(\Delta y_{i,t}) = \mathrm{Var}_{l_{i,t}}(\mathrm{E}[\Delta y_{i,t}|l_{i,t}]) + \mathrm{E}_{l_{i,t}}(\mathrm{Var}(\Delta y_{i,t}|l_{i,t})). \tag{36}$$

Since $[\Delta y_{i,t}|l_{i,t}]$ has mean $l_{i,t}\mu_A$ and variance $l_{i,t}\sigma_A^2$, we have:

$$\mathrm{E}[\Delta y_{i,t}^2|l_{i,t}] = \mathrm{E}[\xi_{i,t}^2(\sum_{j=0}^{l_{i,t}-1} \nu_{i,t-j})^2] = \lambda(l_{i,t}(\sigma_A^2+\sigma_I^2) + l_{i,t}\mu_A^2.$$

A similar calculation shows that:

$$\mathrm{E}[\Delta y_{i,t}|l_{i,t}] = \lambda l_{i,t}\mu_A.$$

Hence:

$$\mathrm{Var}(\Delta y_{i,t}|l_{i,t}) = \lambda l_{i,t}(\sigma_A^2+\sigma_I^2) + \lambda(1-\lambda)l_{i,t}^2\mu_A^2$$

and taking expectation with respect to $l_{i,t}$ (and using the expressions in Lemma B1) leads to:

$$\mathrm{E}_{l_{i,t}}\mathrm{Var}(\Delta y_{i,t}|l_{i,t}) = \sigma_A^2 + \sigma_I^2 + \frac{(1-\lambda)(2-\lambda)}{\lambda}\mu_A^2. \tag{37}$$

An analogous (and considerably simpler) calculation shows that:

$$\mathrm{Var}_{l_{i,t}}(\mathrm{E}[\Delta y_{i,t}|l_{i,t}]) = (1-\lambda)\mu_A^2. \tag{38}$$

The proof concludes by substituting (37) and (38) in (36). ∎

Lemma B5 *For* $N \geq 1$*:*

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

Proof The case $N = 1$ corresponds to Lemma B4. For $N \geq 2$ we have:

$$\mathrm{Var}(\Delta y_t^N) \;=\; \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}(\Delta y_{i,t}, \Delta y_{j,t})$$

$$= \frac{1}{N^2}\{N\text{Var}(\Delta y_{1,t}) + N(N-1)\text{Cov}(\Delta y_{1,t}, \Delta y_{2,t})\}.$$

The first identity follows from the bilinearity of the covariance operator, the second from the fact that $\text{Var}(\Delta y_{i,t})$ does not depend on i and $\text{Cov}(\Delta y_{i,t}, \Delta y_{j,t})$, $i \neq j$, does not depend on i or j.

The remainder of the proof follows from using the expressions derived in Lemmas B3 and B4. ∎

Lemma B6 *Recall that in the main text we defined* $\Delta y_t^* \equiv \sum_{i=1}^N \Delta y_{i,t}^*/N$. *We then have:*

$$\begin{aligned} \text{Var}(\Delta y_t^*) &= \sigma_A^2 + \frac{\sigma_I^2}{N}, \\ \text{Cov}(\Delta y_t^N, \Delta y_t^*) &= \lambda[\sigma_A^2 + \frac{\sigma_I^2}{N}]. \end{aligned}$$

Proof The proof of the first identity is trivial. To derive the second expression we first note that:

$$\begin{aligned} \text{Cov}(\Delta y_{i,t}, \Delta y_{j,t}^*) &= \text{E}[\xi_t(\sum_{k=0}^{l_{i,t}-1} \Delta y_{i,t-k}^*)\Delta y_{j,t}^*] - \mu_A^2 \\ &= \sum_{k\geq 1} \text{E}[\xi_t(\sum_{k=0}^{l_{i,t}-1} \Delta y_{i,t-k}^*)\Delta y_{j,t}^* \mid l_{i,t} = k, \xi_{i,t} = 1]\lambda^2(1-\lambda)^{k-1} - \mu_A^2 \\ &= \lambda^2 \sum_{k\geq 1}[(\sigma_A^2 + \delta_{i,j}\sigma_I^2 + \mu_A^2) + (k-1)\mu_A^2](1-\lambda)^{k-1} - \mu_A^2 \\ &= \lambda(\sigma_A^2 + \delta_{i,j}\sigma_I^2), \end{aligned}$$

with $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. The expression for $\text{Cov}(\Delta y_t^N, \Delta y_t^*)$ now follows easily. ∎

Theorem B1 Δy_t^N *follows an ARMA(1,1) process:*[22]

$$\Delta y_t - \phi\Delta y_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1},$$

where ε_t *denotes the innovation process and*

$$\begin{aligned} \phi &= 1-\lambda, \\ \theta &= \frac{1}{2}(L - \sqrt{L^2-4}). \end{aligned}$$

with

$$L \equiv \frac{2+\lambda(2-\lambda)(K-1)}{1-\lambda}.$$

Also, as N tends to infinity θ *converges to zero, so that the process for* Δy_t^N *approaches an AR(1) (and we recover Rotemberg's (1987) result).*

We also have:

$$\sigma_s^N = \frac{1}{N}\left\{\frac{\lambda}{2-\lambda}(N-1)\sigma_A^2 - \mu_A^2\right\}(1-\lambda)^s, \qquad s = 1,2,\ldots$$

[22]We have that $\phi = \theta$, so that the process reduces to white noise, if and only if $N = 1 + \frac{2-\lambda}{\lambda}\frac{\mu_A^2}{\sigma_A^2}$.

$$\rho_s^N = \frac{K}{1+K}(1-\lambda)^s; \qquad s = 1,2,3,\ldots$$

where σ_s^N and ρ_s^N denote the s-th order autocovariance and autocorrelation coefficients of Δy_t^N, respectively.[23]

Proof We have:

$$\begin{aligned}\mathrm{Cov}(\Delta y_{t+s}, \Delta y_t) &= \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(\Delta y_{i,t+s}, \Delta y_{j,t}) \\ &= N\mathrm{Var}(\Delta y_{1,t}) + N(N-1)\mathrm{Cov}(\Delta y_{1,t+s}, \Delta y_{2,t}).\end{aligned}$$

Where the justification for both identities is the same as in the proof of Lemma B5. The expression for σ_s^N now follows from Lemmas B3 and B4. The expression for the autocorrelations follow trivially using Lemma B5 and the formula for the autocovariances.

The expressions we derived for the autocovariance function of Δy_t and Theorem 1 in Engel (1984) imply that Δy_t follows an ARMA(1,1) process, with autoregressive coefficient equal to $1-\lambda$. The expression for θ follows from standard method of moments calculations (see, for example, equation (3.4.8) in Box and Jenkins (1976)) and some patient algebra.[24] Finally, some straightforward calculations prove that θ converges to zero as n tends to infinity. ∎

Corollary B1 *Proposition 4 is a direct consequence of the preceding theorem.*

Proof Part (i) follows trivially from Proposition 3 and the fact that both regressors are uncorrelated. To prove (ii) we first note that:

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

From Lemma B6 and the fact that Δy_t^* and Δy_{t-1} are uncorrelated it follows that:

$$\begin{aligned}\mathrm{Cov}(\Delta y_t^N - \Delta y_{t-1}^N\,,\ \Delta y_t^* - \Delta y_{t-1}^N) &= \lambda\left(\sigma_A^2 + \frac{\sigma_I^2}{N}\right) + (1-\rho_1^N)\mathrm{Var}(\Delta y_t^N), \\ \mathrm{Var}(\Delta y_t^* - \Delta y_{t-1}^N) &= \sigma_A^2 + \frac{\sigma_I^2}{N} + \mathrm{Var}(\Delta y_t^N).\end{aligned}$$

The expressions derived earlier in this appendix and some patient algebra complete the proof. ∎

[23] The expressions for $\mathrm{plim}_{T\to\infty}\hat{\lambda}^N$ in Proposition 3 follow from noting that $\hat{\lambda}^N = 1-\rho_1^N$.

[24] A straightforward calculation shows that $L > 2$, so that we do have $|\theta| < 1$.