Online Appendices for 'Refining Stylized Facts from Factor Models of Inflation'

Ferre De Graeve and Karl Walentin

March 2014

A Appendix

Appendix A: Aggregation

Since sectoral price indices are combining price quotes across multiple cities, stores and products, one might expect sales, substitutions and general measurement error to average out at the sectoral level. While there definitely is scope for aggregation to reduce the need for our additional components, there are a number of elements that reduce the tendency of these components to be averaged out at the sector level and at the sampled (monthly) frequency. In what follows, we first discuss aggregation under ideal conditions – uncorrelated homogenous-size price changes. We then discuss and quantify two aspects that decrease the power of aggregation: correlated sales or substitutions and heterogeneity in the size of price changes. Throughout we make the simplifying assumption that all products receive equal weights in the sector-level indices.

The discussion below concerns what fraction of the volatility of product-level sales and substitutions remains at the sector level. But let us start by stating that the dynamics, in particular the persistence properties, induced by these phenomena remain unchanged by aggregation: An *iid* movement induced by substitution at the product level induces an *iid* movement in the corresponding sector index. Similarly for the MA component induced by sales.[1]

The first reason product level measurement errors do not completely cancel out at the sectoral level is that the number of product prices sampled per month is limited. The consumer price index (CPI), which is the main source of the sectoral PCE price indices we use, is based on 70.000-80.000 prices across 388 entry-level items (ELIs) roughly corresponding to the PCE sectors we study, yielding a mean number of observations slightly above 200

[1] Recall eq. (8), which at the sector level yields

$$M_{it} = \sum_j \left(\xi_{jit} - \xi_{jit-1}\right)$$

where j indexes products within a sector and ξ_{jit} is uncorrelated across t. Then $Var(M_{it}) = 2Var(\xi_{ji})$ and autocorrelation at the sector level is

$$\begin{aligned}
\rho\left(M_{it}, M_{it-1}\right) &= \frac{1}{Var(M_{it})} Cov\left(\sum_j \left(\xi_{jit} - \xi_{jit-1}\right), \sum_j \left(\xi_{jit-1} - \xi_{jit-2}\right)\right) \\
&= \frac{1}{Var(M_{it})} Cov\left(\sum_j \left(-\xi_{jit-1}\right), \sum_j \left(\xi_{jit-1}\right)\right) = \\
&= -\frac{Var\left(\sum_j \xi_{jit-1}\right)}{Var(M_{it})} = -0.5
\end{aligned}$$

which coincides with the product-level autocorrelation of M_{jit}.

product prices per ELI/PCE sector and month. Theoretically, in absence of any aggregation problems, the ratio of the standard deviation of the index, σ_{index}, to the standard deviation of the product price, $\sigma_{product}$, is $\frac{1}{\sqrt{N}}$. This implies that for the sector with the mean number of observations $1/\sqrt{200} = 7\%$ of the variation induced by sales and substitutions at the product level would remain at the sector level.[2] The first column in Table A.1 present the corresponding numbers for the empirically relevant range of sample sizes.

Correlated sales or product substitutions could occur due to sector-specific shocks: low demand can build up inventory and induce larger sales, technical progress can generate product turnover and induce product substitutions, etc.[3] To illustrate the impact of correlated sales or substitutions we perform the following exercise. For a sample length equal to ours (T=353) we randomly generate sequences of sales (the outcomes are indistinguishable for the case of substitutions). At any point in time, an individual product is on sale with a particular frequency. If there is no sale, the price remains constant. When there is a sale, the price change is a sum of two random components from the normal distribution: A common component generates correlated variation across products within an index and an idiosyncratic component generates uncorrelated variation. We generate many product level price series, and construct inflation indices from them, for a variety of numbers of goods in the index, N. In this exercise the only reason that the theoretical prediction of the effect of aggregation, $1/\sqrt{N}$, does not obtain is that the size of sales contain a common component that makes them correlated. We let the correlation equal 0.25. In Table A.1 we present the results for a range of frequencies, recalling from Section 4 that micro evidence indicates that the median monthly frequency of sales are in the range from 7.4% to over 20%, and 3.4% to 5% for item substitutions. The first, and least surprising, result to note is that correlated sales do not average out very well. Secondly, aggregation actually works better the lower the frequency is. The intuition is that for low frequencies the realized correlation tends towards zero as most prices are unchanged. To specifically address the question of how well aggregation works for the median sector, we read from the table that for $N = 200$, the ratio of the standard deviation of the index relative to the standard deviation of its underlying products $\frac{\hat{\sigma}_{index}}{\sigma_{product}}$ is roughly 0.2 at the empirical frequency of sales and roughly 0.1 at the empirical frequency of substitutions. Interestingly, results at the empirical frequency of sales are approximately unchanged for $N = 500$ and $N = 1000$. In other words, roughly 20% (10%) of the product

[2] Whether that 7% represents a large fraction of the index's variance, which also contains regular price changes, is a different question. It depends on the relative volatility of sales and substititions vs. regular price changes at the product level. Micro level data suggest that sales and to a smaller degree, substitutions, may well cause substantially more volatility than regular price changes (see Section 4 for details). This makes effectively controlling for them at the index level all the more needed.

[3] Note that the price data we work with is seasonally adjusted, so correlation in sales that follow a seasonal pattern are filtered out.

level volatility from sales (substitutions) remains at the sector level if correlation is 0.25. This is substantially more than for uncorrelated price changes.

Table A.1: Aggregation and sales/substitutions - correlation

		Frequency			
Number of products in index: N	$1/\sqrt{N}$	0.25	0.1	0.05	0.01
50	0.1414	0.2849	0.2110	0.1796	0.1495
100	0.1000	0.2685	0.1865	0.1497	0.1113
200	0.0707	0.2595	0.1728	0.1319	0.0864
500	0.0447	0.2536	0.1640	0.1205	0.0670
1000	0.0316	0.2519	0.1611	0.1162	0.0591

Note: The table reports the ratio of the standard deviation of an index, $\hat{\sigma}_{index}$, relative to the (homogenous) standard deviation of its underlying products, $\sigma_{product}$, for various N and frequencies, but for a fixed correlation of 0.25. The first column is the theoretical relation without correlation and the four subsequent columns the small-sample (T=353) results across 5000 replications.

It is plausible that not all products within a sector exhibit the same unconditional size of sales or substitutions. Heterogeneity in size of sales or substitutions within a sector weakens aggregation. Intuitively, the degree to which various sales or substitutions cancel out at the sector level decreases with size heterogeneity.

To quantify the effect of heterogeneity we perform a similar exercise to the one above. We let the size of the sale or substitution be a random draw from a normal distribution whose standard deviation is drawn from a uniform distribution to induce heterogeneity in size. As a rough reference for the within-sector size heterogeneity we use heterogeneity between major groups from Nakamura and Steinsson (2008). It shows that the standard deviation of the sales size, σ_{size}, is one third of the mean sales size, μ_{size}, for both of the sample periods they report.

We report the results for a range of heterogeneity in Table A.2. We note that the quantitative impact of heterogeneity in size is limited for this range of heterogeneity. Results are indistinguishable for sales and substitutions, and independent of frequency.

In this section we have quantified how much of product-level variation in prices due to sales and substitutions remains at the sector-level. We first noted that the empirical sample size in the mean sector is limited. This makes it likely that sales and substitutions generate significant variance at the sectoral index level. We then separately quantified the impact of two factors that further weaken aggregation: correlation and heterogeneity. Empirically, across sectors, there are different numbers of products per sector, varying degrees of heterogeneity across products within each sector, and varying degrees of correlation between those products. Each of these factors, and possible interactions between them affect how well aggregation works.

Table A.2: Aggregation and sales/substitutions - heterogeneity

		σ_{size}/μ_{size}				
Number of products in index: N	$1/\sqrt{N}$	0.95	0.75	0.5	0.25	0.05
50	0.1414	0.1952	0.1761	0.1577	0.1456	0.1416
100	0.1000	0.1376	0.1247	0.1118	0.1031	0.1000
200	0.0707	0.0973	0.0882	0.0790	0.0728	0.0707
500	0.0447	0.0616	0.0558	0.0500	0.0460	0.0447
1000	0.0316	0.0436	0.0395	0.0353	0.0325	0.0317

Note: The table reports the ratio of the standard deviation of an index, $\hat{\sigma}_{index}$, relative to the mean of the heterogenous standard deviation of its underlying products, $\sigma_{product}$, for various ratios of the within sector standard deviation of the size of sales, σ_{size}, to the mean size of sales, μ_{size}. The first column is the theoretical relation without heterogeneity, the four subsequent columns the small-sample (T=353) results for lower frequencies of price change across 5000 replications.

Appendix B: Estimator properties in finite samples of simulated data

This appendix documents empirical properties of the maximum likelihood estimator used in the paper. We also quantify the bias from estimating an AR process when the DGP consists of multiple components. In particular, we simulate data from various one- and multicomponent processes for sample lengths equal to our data ($T = 353$). For each of these, we estimate single component (P, as in eq. (4), henceforth AR) and multicomponent processes ($P + I + M$, as in eq. (5)-(8), henceforth PIM). For each process we use one lag for the AR (P) component. The Monte Carlo results are based on 100 time series per data-generating process. The data is generated from

$$e_t = P_t + I_t + M_t$$

with

$$\begin{aligned} P_t &= \rho P_{t-1} + \varepsilon_t \\ I_t &= \epsilon_t \\ M_t &= \xi_t - \xi_{t-1} \end{aligned}$$

for the parameter values in Table A.3.

Consider the last column of Table A.3, PIM high. Here all three components are equally important, and the persistent component is very persistent. Figure A.1 shows how, even for data with a limited time dimension, the estimator has no problem disentangling the various components.

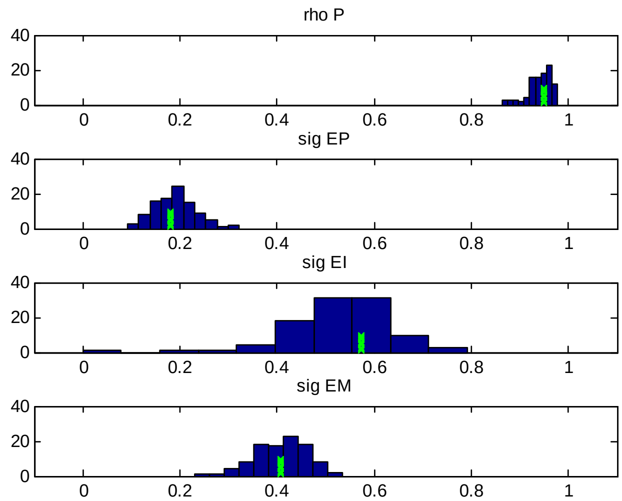
It is plausible that high persistence makes identification easier. Therefore, now consider a PIM process with intermediate persistence, PIM low in Table A.3. In this case, as ap-

Table A.3: Data generating processes for artificial data

	IID	AR low	AR high	PIM low	PIM high
ρ	0	0.5	0.95	0.5	0.95
σ_P^2	1	1	1	.33	.33
σ_I^2	0	0	0	.33	.33
σ_M^2	0	0	0	.33	.33

Note: To facilitate evaluation of the relative importance of the various components, the table specifies volatility of the components rather than the innovations. Thus, $\sigma_P^2 = \frac{\sigma_\varepsilon^2}{1-\rho^2}$, $\sigma_I^2 = \sigma_\epsilon^2$, $\sigma_M^2 = 2\sigma_\xi^2$ and the three shocks are orthogonal and follow $(\varepsilon_t, \epsilon_t, \xi_t)' \sim N(0_{3\times1}, D)$.

Figure A.1: Estimation on simulated data: PIM on PIM high

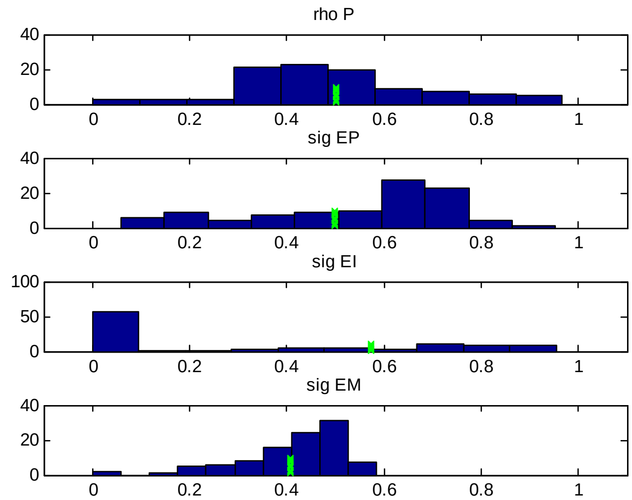


Note: Green x's mark data-generating parameters

parent from Figure A.2, there is more dispersion in point estimates. Persistence tends to be slightly underestimated (and, accordingly, the volatility of the persistent shock slightly overestimated). The M component is still consistently identified, while the I component is not always easily detected.

Now consider the alternative; estimating an AR specification on these data. Irrespective of the persistence of the underlying process, estimating an AR fails to detect any significant amount of persistence, as illustrated in Figure A.3 and Figure A.4. We interpret these simulations as follows. While for low-persistence multicomponent processes, PIM-specifications may imply substantial imprecision regarding the variances of the components, they allow a fairly adequate evaluation of persistence. When persistence is high, they are both unbiased and precise across repeated samples, for the empirically relevant sample lengths. For the same DGP's, AR-specifications are clearly inadequate. These simulations establish one type

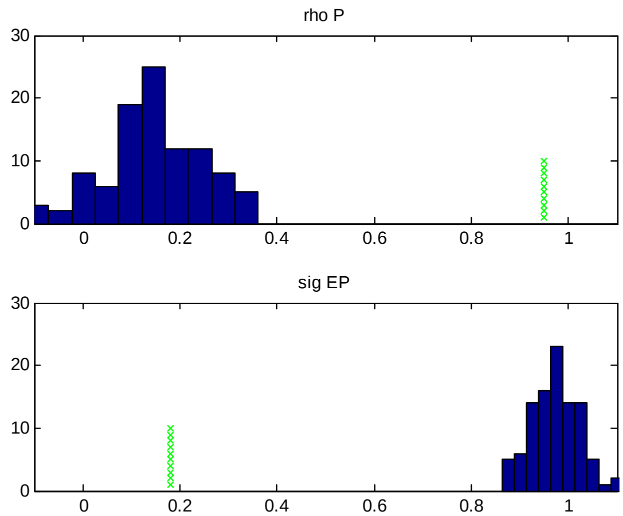
Figure A.2: Estimation on simulated data. PIM on PIM low



of risk: if the DGP is a multicomponent process, AR estimation will fail to detect persistence.

The question remains as to how PIM-specifications perform in the case of AR-DGPs. It is possible that the cure is worse than the disease. Figure A.5 shows that this type of risk is limited. In particular, for an AR-DGP with high underlying persistence estimating a PIM-specification comes at little cost. As persistence decreases, see Figure A.6, PIM-estimation attributes some variation to the I component, which entails a minor overestimation of persistence. Taken to the limit, estimating PIM-specifications on *iid* data, as in Figure A.7, identification of separate components is cumbersome: there is a lot of dispersion in all the estimates. Firstly, however, note that the modes of the distributions are typically located at the truth. Secondly, for persistence close to zero, the likelihood is flat in certain dimensions: this occurs as P and I become equivalent.

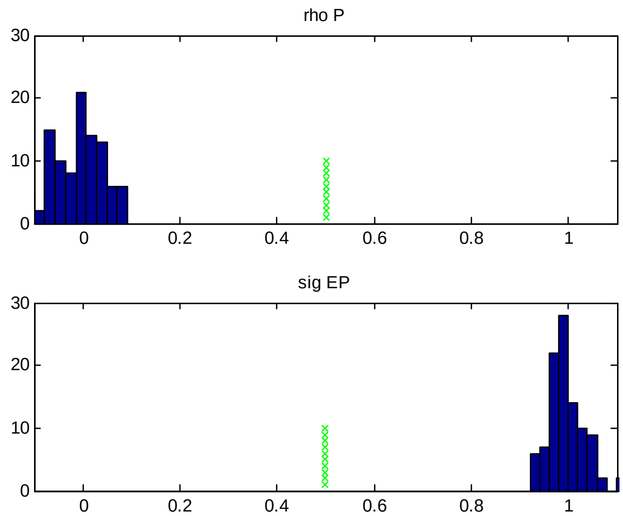
Figure A.3: Estimation on simulated data. AR on PIM high



Appendix C: Comparison of factor loadings - benchmark vs. simple

Figure A.8 compares the estimated loadings for 190 PCE sectors on common factors of the benchmark model (with 3 sectoral components) and the simple model (with one single sectoral component). Correlations are 0.99 except for the last factor with correlation 0.97.

Figure A.4: Estimation on simulated data. AR on PIM low

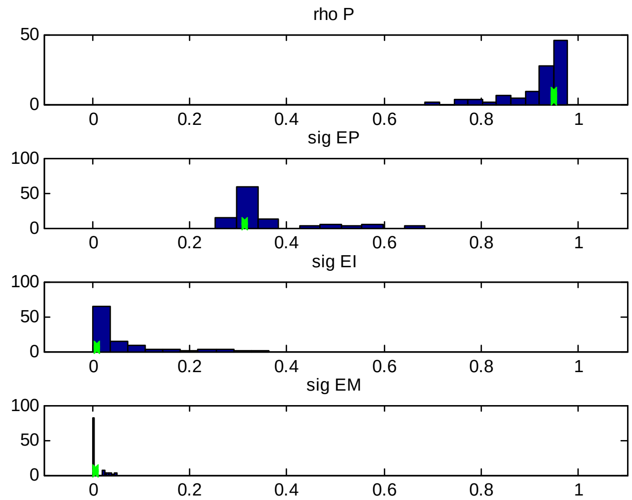


Appendix D: Robustness of results

The main results of the benchmark model go through for other data sets and for variations in the model specification considered. First, as in Boivin et al. (2009), we consider the effect of shortening the sample period to 1984-2005. This serves to isolate the results from the very different behavior of macroeconomic aggregates prior to and during the early eighties disinflation and the start of the so-called Great Moderation. Figures A.9 and A.10 document the variance and persistence of the various components for this period. Compared to the full sample results documented in Figure 6 the relative variance of aggregate shocks is substantially smaller already in the simple model. This is not unexpected, since decreased variance of aggregate conditions is exactly what the Great Moderation represents. Comparing the relative importance of aggregate shocks in the simple factor model with that of the benchmark model, which accounts for sales and substitutions, again shows how the former model substantially overestimates the relative importance of the sector-specific component. While the traditional approach suggests that in the median sector idiosyncratic shocks are roughly 14 times more important than aggregate shocks, the benchmark model finds this to be only 6 times as large. One could argue that this high relative variance of idiosyncratic shocks was particular to the Great Moderation era and might well disappear when considering more recent data.[4] Turning to persistence in Figure A.10, the results for the subsample

[4]Unfortunately, a change in the PCE definition makes extending the sample and verifying this conjecture infeasible.

Figure A.5: Estimation on simulated data. PIM on AR high



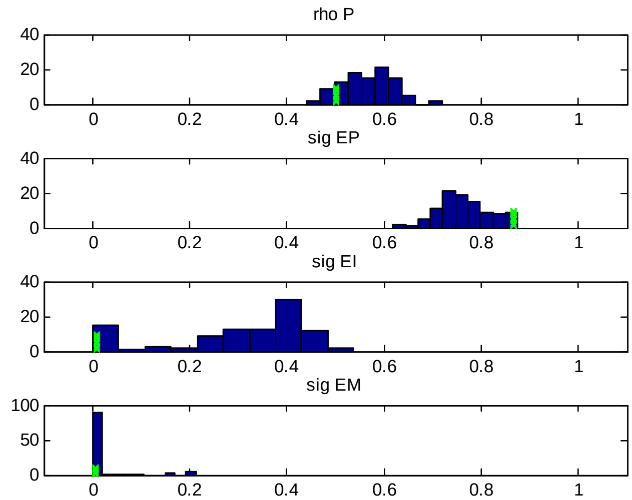
are very similar to those for the full sample. A simple factor model reveals no persistence due to sectoral shocks for the median sector, while substantial persistence is visible in the model that accounts for measurement error, sales and substitutions. Again, one observes the strong concentration of sectors at very high levels of persistence.

Second, to assess the generality of their results, Boivin et al. (2009) also consider sectoral PPI series, and document that the stylized facts continue to hold. As an additional robustness check, we therefore re-estimate the simple model and the benchmark factor model for the PPI data. Here too, the results are very similar: The simple model confirms the first stylized fact and estimates sectoral shocks to be 9 times more volatile than aggregate shocks for the median sector (Figure A.11). The benchmark model reduces this ratio to below 4. In terms of persistence, too, a similar bias appears to be present. As is clear from Figure A.12, the standard, simple approach finds no persistence – stylized fact *(ii)* – while the benchmark approach indicates substantial persistence.[5]

Third, we now switch from documenting robustness in terms of data to robustness in terms of model specification. Recall that our sales definition, operationalized by eq. (8), is the most restrictive among the alternatives in the literature, possibly not capturing all sales in the data. We therefore also explore a less restrictive sales definition that replaces eq. (8)

[5]Micro price studies show that sales are rather uncommon in producer prices. On the one hand, this may reduce the likelihood of the M-component to capture sales, but rather e.g. measurement error. On the other hand, a lower incidence of sales at the micro-level can also reduce the likelihood of them averaging out at the sector-level, in which case M would absorb sales.

Figure A.6: Estimation on simulated data. PIM on AR low



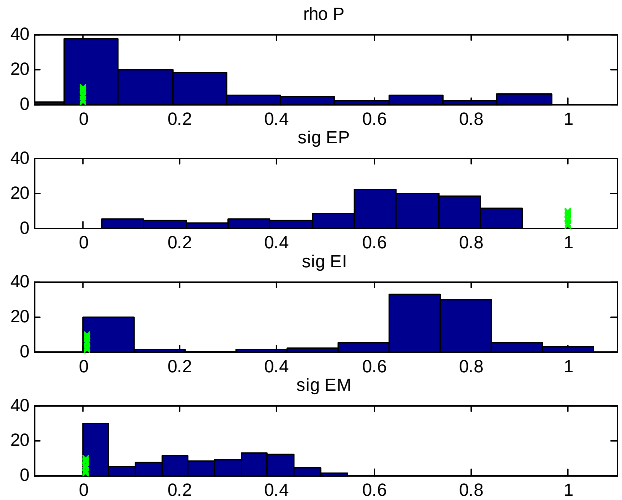
by

$$M_{it} = \rho_{m,i}(L)M_{it-1} + \xi_{it}$$

and where identification is achieved by restricting the sum of the lags to be negative, $\rho_{m,i}(1) < 0$, while for the persistent component, P_{it}, we require $\rho_i(1) > 0$. Also this alternative specification yields very similar results to our benchmark model, both in terms of volatility of each component and persistence of P_{it}.

Finally, we perform a robustness exercise where we reduce the lag length of the persistent component, P_{it}. The reason for this exercise is that 13 lags may over-parameterize the model, in particular in the presence of the two additional components. The results are very similar to our benchmark specification when either imposing 3 lags or using standard lag selection criteria.

Figure A.7: Estimation on simulated data. PIM on iid



Appendix E: Isolating measurement error using quantities

The observation equation for sector i becomes

$$\pi_{it} = \lambda_i^{\pi\prime} C_t + P_{it} + I_{it} + M_{it} + \eta_{it} \tag{A.1}$$

$$q_{it} = \lambda_i^{q\prime} C_t + \alpha_i^P P_{it} + \alpha_i^I I_{it} + \alpha_i^M M_{it} + \varsigma_{it}. \tag{A.2}$$

or

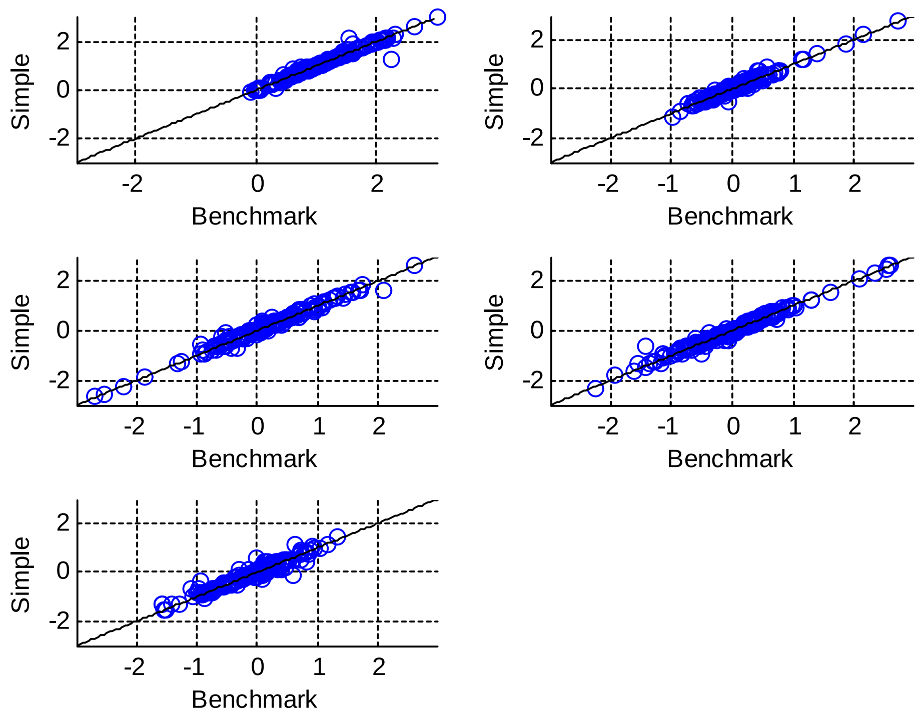
$$\begin{bmatrix} \pi_{it} \\ q_{it} \end{bmatrix} = \begin{bmatrix} \lambda_i^{\pi\prime} \\ \lambda_i^{q\prime} \end{bmatrix} C_t + \begin{bmatrix} 1 & 1 & 1 \\ \alpha_i^P & \alpha_i^I & \alpha_i^M \end{bmatrix} \begin{bmatrix} P_{it} \\ I_{it} \\ M_{it} \end{bmatrix} + \begin{bmatrix} \eta_{it} \\ \varsigma_{it} \end{bmatrix}$$

Here q denotes quantity growth. In addition to the requirement that the three components P, I and M affect quantities, their persistence properties continue to hold, as in eqs. (6)-(8). Measurement error in inflation and quantity growth are denoted by η_{it} and ς_{it} respectively. They are identified because they affect price or quantity respectively, but not both.

In the PCE data used by Boivin et al. (2009) real quantities are available, as part of X_t. However, real quantities are not measured independently, but calculated as nominal quantity deflated by the price index. To ensure that measurement error does not affect the quantity variable we therefore use nominal quantities.

In eq. (A.1), as before, the I and M components absorb substitutions and sales, respectively. The importance of measurement error is now captured separately by the sector-specific

Figure A.8: Loadings on the 5 common factors.



component η_{it}. Note that substitutions related to sampling (a product not being available at the surveyed retailer) will not be captured by the I component in this setting, but instead by the measurement error component for inflation, η_{it}.

We allow both the idiosyncratic inflation and quantity components η_{it} and ς_{it} to exhibit unrestricted autoregressive dynamics. The reason for this flexible specification is that, for the inflation equation, for instance, measurement error in prices would generate negative autocorrelation.

Note that the identification assumption that the P, I and M components affect quantities does not hold at $\alpha_i = 0$. This case does not turn out to be practically important.

The last two columns of Table 5 in the main text summarize the results of estimating (A.1)-(A.2), subject to (6)-(8). The following figures show the results for the relative variance (Figure A.13) and persistence (Figure A.14). They are very similar to the results of the benchmark factor model presented in the main text.

Figure A.9: Variance - subsample

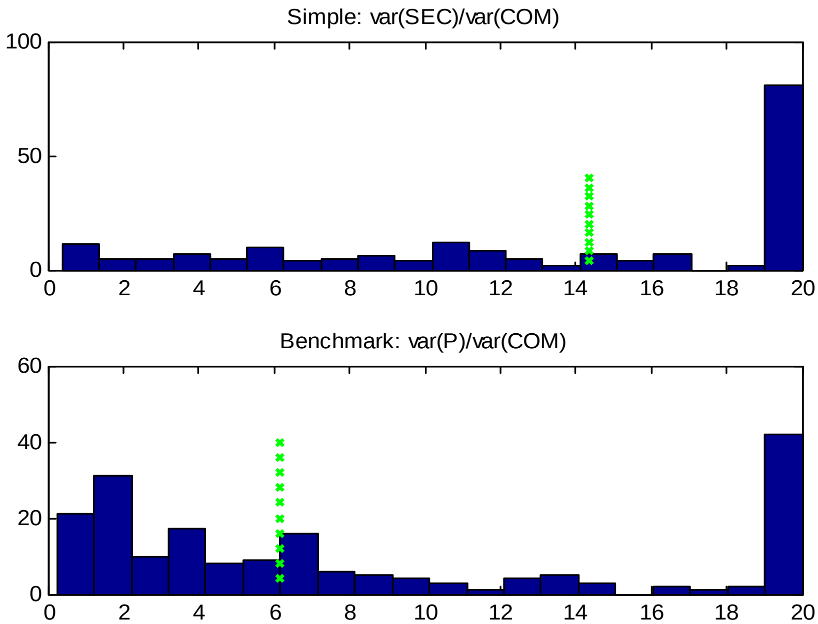


Figure A.10: Persistence - subsample

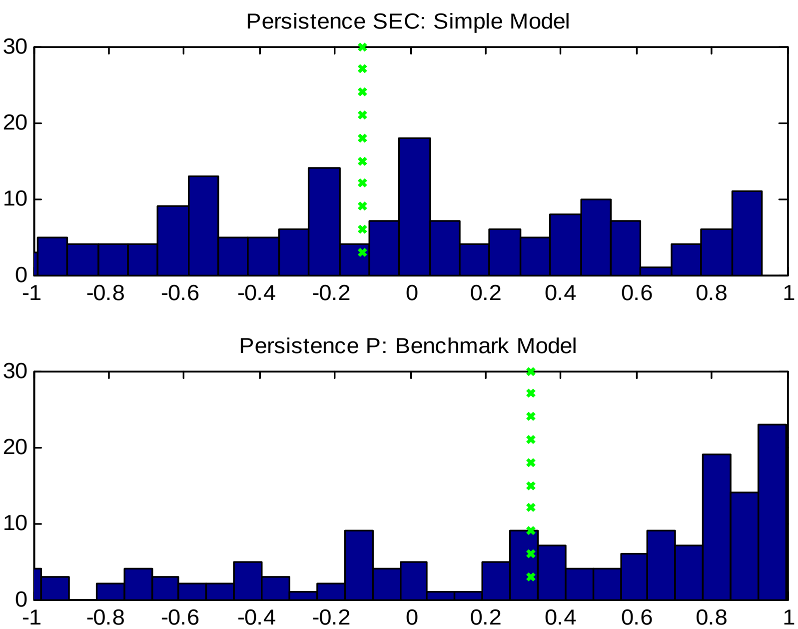


Figure A.11: Variance - PPI

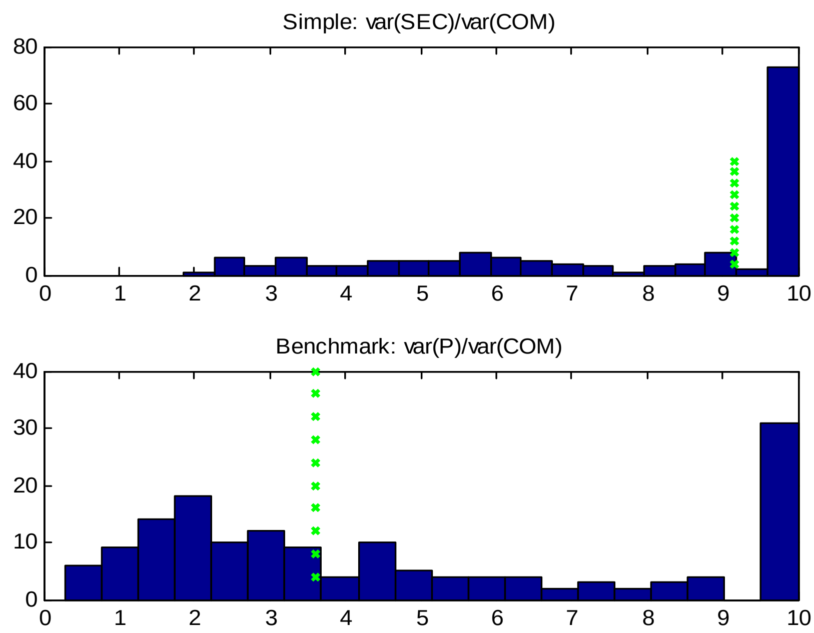


Figure A.12: Persistence - PPI

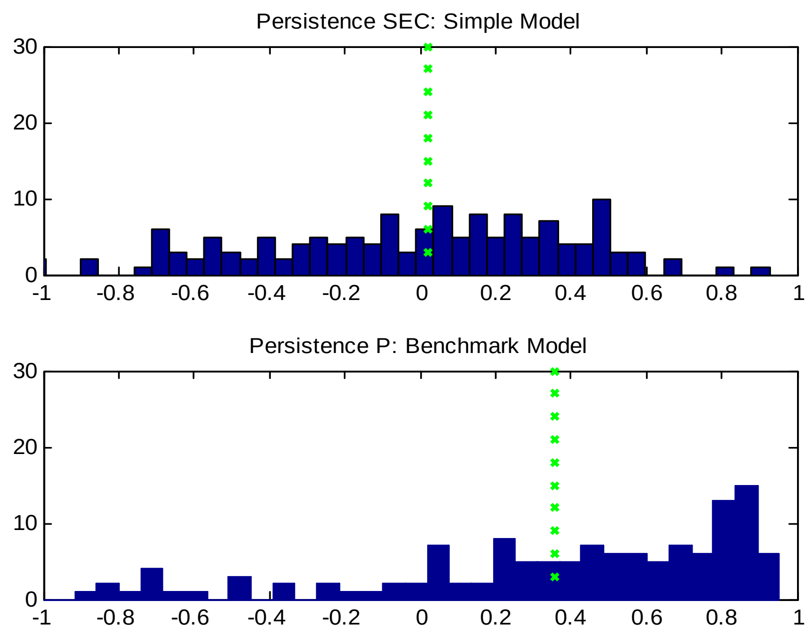


Figure A.13: Identification using quantities - variance

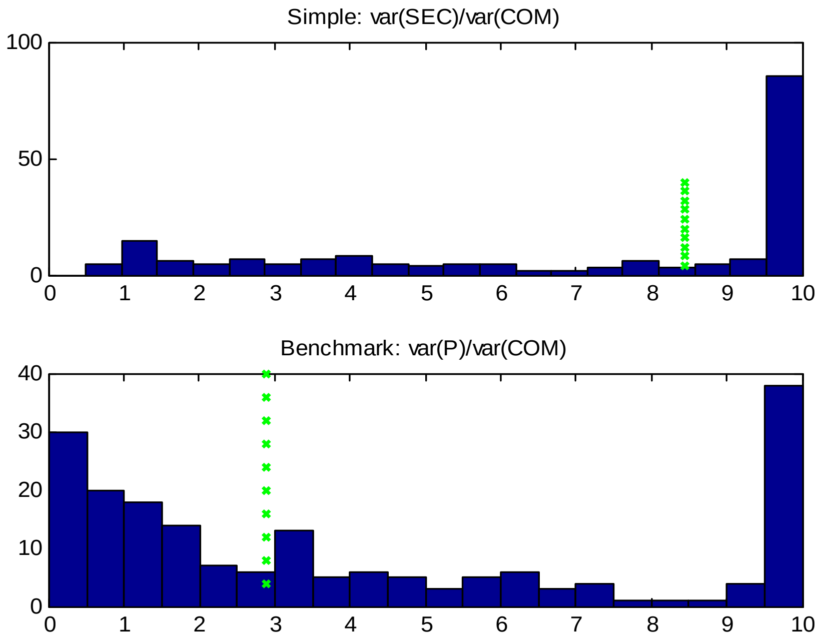


Figure A.14: Identification using quantities - persistence