

Labor Market Shortages and the Phillips Curve*

Identification under Heterogeneous Exposure to Common Shocks

[Updated regularly – latest version [here](#).]

Cristian Arcidiacono[†]

2026-10-11

Abstract

Cross-sectional designs with time fixed effects are widely used in empirical macroeconomics, including Phillips curve estimation. We show that heterogeneous exposure to common shocks can bias slope estimates, conflate responses to idiosyncratic and synchronized slack, and overstate precision. We propose a two-step extension of interactive fixed effects that separates these responses under testable conditions. Using firms' reported labor shortages across euro area sectors and countries, we find that a one standard deviation increase in a sector's own tightness raises its inflation by about 0.25 percentage points – three times the time fixed effect estimate. A one standard deviation increase in synchronized tightness raises it by about 1.3 points, revealing a substantially stronger inflation response when shortages are pervasive.

Keywords: Phillips curve; interactive fixed effects; labor market tightness; common shocks; heterogeneous exposure; euro area.

JEL Codes: C23; C38; E24; E31; J63.

*I am deeply indebted to Pierpaolo Benigno and Ricardo Reis for their support all along. I thank Matthias Gnewuch, Douglas Araujo, Simone Emiliozzi, Elisa Guglielminetti, Dante Gaviano, Mathias Klein, Joe Hazell, Anastasia Burya, Giulia Gitti, Efrem Castelnuovo and various seminar participants for helpful comments and discussions. I also thank Gabriele Iannizzi for insightful discussions and for his help at the early stage of the project. The paper received the INFER Best PhD Paper Award at the 28th INFER Annual Conference. Financial support from INFER is gratefully acknowledged.

[†]University of Bern, cristian.arcidiacono@unibe.ch.

1 Introduction

The Phillips curve is among the most studied objects in macroeconomics. Ever since [Phillips \(1958\)](#), economists have tied low unemployment to inflation and yet, seventy years on, its consistent empirical estimation remains elusive. Recent work has emphasized several challenges, with four questions open: whether the curve is flat ([Hazell et al., 2022](#); [McLeay and Tenreyro, 2020](#); [Furlanetto and Lepetit, 2025](#)); whether its slope is constant, or steepens once labor markets tighten ([Benigno and Eggertsson, 2023](#); [Gitti, 2024](#); [Hooper et al., 2020](#)); how inflation expectations and supply shocks should be controlled for ([Reis, 2026](#); [Beaudry et al., 2025](#)); and how slack should be measured ([Albuquerque and Baumann, 2017](#); [Hong et al., 2018](#)). This paper’s identification strategy allows units to respond differently to common shocks. Conventional time fixed effects impose a common response, overlooking heterogeneity that can bias the estimated Phillips curve slope and conflate two conceptually distinct components of the response: the effect of unit-specific labor market tightness and the effect of synchronized tightness operating across units.

In the last decade, the estimation of the Phillips curve has increasingly adopted cross-sectional designs. Where aggregate time series often left the curve only weakly identified ([Mavroeidis et al., 2014](#); [McLeay and Tenreyro, 2020](#)), cross-sectional data provide a richer source of identifying variation¹. The appeal of this design is twofold. First, inflation expectations and aggregate supply shocks affect all units simultaneously, so a full set of time fixed effects can absorb these forces without the need to measure them directly. Second, the cross-section provides a source of identifying variation that aggregate data cannot offer. At any given moment, some local labor markets are tighter, and others looser, than the economy as a whole. Once common confounders are accounted for, this cross-sectional variation can identify the slope of the Phillips curve and whether it changes with labor market tightness.

Three identification problems with time fixed effects. However, this paper shows that the strategy becomes problematic in three distinct ways in the setting that motivates it. The first failure is a *biased slope*. Inflation in a given unit responds to common forces in ways that are specific to that unit and correlated with how its slack moves: when energy costs jump for the euro area, producer prices in, say, German chemicals – an industry that burns gas as fuel and feeds on it as raw material – respond quickly and almost fully, while prices in, say, Portuguese apparel barely move – and the units whose prices answer the common shock most strongly may also be the same ones whose hiring difficulties swing most with

¹The literature has extended this approach to increasingly granular data. [Gagliardone et al. \(2025\)](#), for instance, exploit variation at the firm–industry–quarter level for Belgian data.

the common cycle. A period dummy removes only the average response to such a force; each unit’s deviation from that average survives, in the regressor and in the residual alike, and their correlation is what biases the slope. The second failure is that the coefficient is a *blend of two distinct responses*. Measured slack has a common cycle of its own, and inflation need not respond to a tightening shared by everyone as it does to one unit’s own: inflation expectations, wage norms and the willingness to reprice answer economy-wide conditions, not the hiring troubles of a single unit. Both responses matter – the slope on a unit’s own tightness is the micro-local object cross-sectional designs exist to recover; the slope on synchronized tightness is the aggregate one, which captures the shared dimension of the response. When the two responses differ, an estimator that imposes a single slope returns a weighted average of them, with the weight determined by the share of synchronized variation left in the regressor after absorption. The third problem is that, even when the point estimate is centered correctly, its *reported precision may be overstated*. The dummies remove the average inflation response to the common cycle, but not the differences around that average. Two decades of quarterly data may look like tens of thousands of observations, but they are repeated readings on the same few hundred sector-specific sensitivities, taken along a small number of aggregate factors. Standard errors that pool over time treat these repeated observations as if they contained new independent news, so the reported interval keeps shrinking even though uncertainty from the common cycle does not; more data can then buy false confidence.

A general framework. These findings apply beyond the Phillips curve. They hold *whenever units differ in their exposure to common shocks*, and such differential exposure to an aggregate is common to the fiscal, monetary, trade and wealth-effect literatures². The variation these designs use to identify is the variation that breaks the estimator. We therefore state and prove the results in Section 3 for a general class of models, in a framework where a latent common factor enters the outcome and the regressor with unit-specific loadings. We show that two properties of the data play distinct roles in what follows. First, the distortions arise under *heterogeneity*: units respond differently to the same aggregate shock, which a period dummy cannot control for. Second, the solution relies on *granularity*: no unit is large enough to move the aggregates it is exposed to³. Fine data deliver both at once, so the

²Some notable examples include, but are not limited to, Nakamura and Steinsson (2014); Pennings (2021); Ottonello and Winberry (2020); Autor et al. (2013); Chodorow-Reich et al. (2021).

³Where Gabaix (2011) showed that concentration in the size distribution can make idiosyncratic shocks the origin of aggregate fluctuations, we show that its absence makes idiosyncratic variation a clean instrument for partial-equilibrium identification – the same Herfindahl, read in the opposite direction – and we treat his lesson, that this is settled by the size distribution rather than by the number of units, as the discipline our identifying assumption must meet.

same disaggregation that creates the problems supplies its solution; but heterogeneity and granularity play distinct roles and should be assessed separately.

The repair. If these problems arise because units respond differently to common factors, the fix is to model rather than difference them away⁴: interactive fixed effects (Bai, 2009) impose a low-rank structure on unobserved drivers of inflation – inflation expectations, energy prices and common cost-push forces – while allowing each unit to have its own exposure to the factors. A period dummy is the special case in which all units respond identically to the common factor. Freeing them estimates and removes the unit-specific component left in the residual, recentering the slope and correcting its standard error. We show when this works, including under general-equilibrium feedback from units to aggregates, where the granularity condition above is sufficient.

What the estimator requires is a well-measured, unit-level measure of slack. Aggregate inflation expectations, supply shocks, the policy stance and each unit’s exposure to them remain latent and are estimated. Instead, slack is the one object that must be brought to the data. IFE relies more on idiosyncratic slack variation, making it sensitive to measurement error. In our simulations, its advantage over time fixed effects disappears when noise accounts for roughly half the regressor’s variance, as the estimated factors increasingly absorb the underlying signal.

A granular measure of slack. Therefore, the second contribution of this paper is a new measure: the *Perception of Labor Shortages* (PLS) indicator, constructed from the European Commission’s harmonized Business and Consumer Surveys. It records the share of firms reporting labor shortages as a factor limiting their production. It is a question put to nearly thirty-eight thousand manufacturing firms across the EU survey, quarterly since 1990, and available for 24 sub-sectors in all euro area countries. The unemployment rate, the output gap and similar measures describe the average state of the labor market⁵, whereas what drives real marginal cost, and hence inflation, in New Keynesian models with search and matching frictions is the difficulty of hiring the *marginal* worker (e.g., Pissarides, 2000). The PLS indicator asks firms about that difficulty. Moreover, its availability at a disaggregated level is not merely convenient: it supplies the granularity on which the estimator’s consistency rests. Section 6 sets out the construction and the evidence that PLS is accurate in the sense the estimator requires.

⁴Notice that an interesting aspect of this approach is that it naturally relates to the literature using principal components to extract common inflation dynamics across countries, such as Ciccarelli and Mojon (2010) and Auer et al. (2025).

⁵Barnichon and Shapiro (2024); Albuquerque and Baumann (2017).

The euro area Phillips curve, re-estimated. Armed with the measure and the method, we re-estimate the euro area Phillips curve. We hold the exercise to the demanding case the recent literature insists on – specifications that control for short-term inflation expectations (Reis, 2026; Beaudry et al., 2025) – and compare the time fixed effect estimate with the micro-local slope with IFE. At the micro-local slope we recover, a one standard deviation increase in a sector’s own labor-market tightness raises producer-price inflation by roughly a quarter of a point – about three times the response in the time fixed effect design. Our decomposition also shows why. Once factor exposures are estimated, the competing explanations for the gap can be quantified: the omitted variable bias emphasized in the literature is negligible. The gap instead reflects a change in the estimand, together with the co-movement of inflation with euro area-wide tightness that the period dummies delete. Period dummies do not fully purge synchronized movements; they force them into a common coefficient while leaving part of that movement in the regressor. The factor estimator, instead, retains this synchronized component – the part that contains the aggregate content of the response, and whose removal from cross-sectional designs is central to the missing-intercept literature (Wolf, 2023; Moll and Hanney, 2025). An instrument constructed from within the panel, following Sarto (2025), isolates exogenous variation in synchronized tightness and implies a causal response about four times as large as the micro-local slope, and a one standard deviation increase in synchronized tightness raises inflation by about 1.3 percentage points.

Implications. With the Phillips curve re-estimated, the four questions posed at the outset can be answered. How steep is it? There is no single slope: inflation responds modestly to tightness confined to one sector, but much more strongly when tightness is synchronized across the economy; time fixed effects largely remove the variation that reveals this difference. Is it non-linear? The nonlinearity is concentrated in synchronized tightness: its slope is close to zero in normal times but rises sharply in tight aggregate states, while the micro-local slope changes little. How should common inflation expectations and supply shocks be treated? By modelling their heterogeneous transmission rather than absorbing them with common time effects. And how should slack be measured? At the margin: in models with hiring frictions, inflation responds to the cost of the marginal hire, and PLS measures precisely that difficulty while providing the sub-sector-country granularity the estimator requires.

These findings carry implications at two levels. The first is specific to monetary policy. Since firms’ perceptions of labor shortages proxy well for slack, our results underscore the value of reported hiring difficulties as an early warning of inflationary pressure: several central banks

already consult qualitative surveys of this kind⁶, and a systematic, sector-disaggregated reading would sharpen the monitoring of inflation risk. The non-linearity we document adds the sharper message: because inflation becomes far more sensitive to shortages once the labor market is tight, linear models understate inflation risk precisely when labor markets are tight and overstate the output cost of disinflation – the two moments when the slope of the curve matters most.

The second is methodological. Granular panel data are increasingly the norm in applied macroeconomics – sub-sector by country, firm by market, region by industry – and our analysis shows that absorbing common shocks with time dummies becomes *more* hazardous, not less, as the data become finer: richer cross-sections carry richer heterogeneity. Any literature that reads a structural slope off granular variation is exposed, and the correction we propose requires no data beyond the panel already in hand.

The paper is organized as follows. Section 2 situates the contribution within the related literature, with a detailed comparison in the appendix. Section 3 sets out the general model, defines the micro-local estimand and characterizes the three failures of time fixed effects. Section 4 states the assumptions under which IFE recovers the micro-local slope and restores valid inference. Section 5 presents Monte Carlo evidence on the three failures and the IFE correction. Section 6 develops the PLS measure of labor market tightness and the sense in which it satisfies the requirement the theory imposes. Section 7 re-estimates the New Keynesian Phillips curve, covering data, estimation and empirical findings; it then documents the non-linearity of the inflation–slack relationship (Section 7.5). Section 8 concludes.

2 Related Literature

The paper sits at the intersection of three literatures. On the *application*, we do not aim to survey the large and ever-growing literature on the Phillips curve⁷; we focus on those papers that are more closely related to ours. The paper enters the literature that recovers the Phillips curve slope from cross-sectional and panel variation rather than from aggregate time series (Fitzgerald et al., 2024; Hazell et al., 2022; McLeay and Tenreyro, 2020) – a strand distinct from the aggregate NKPC identification literature, for which Mavroeidis et al. (2014) document a weak identification of the slope in time-series data – and extends it to the euro area, where estimates have remained aggregate or country-level⁸. We introduce a granular,

⁶For example, the U.S. Federal Reserve’s Beige Book.

⁷For a recent review see Furlanetto and Lepetit (2025).

⁸Mostly with country-level proxies, except Levy (2019) and Beschin et al. (2025) who used NUTS-2 data, which however is limited to annual frequency.

sub-sector-country level measure of labor market tightness⁹ and show that much of the apparent flatness of the euro area Phillips curve reflects what time fixed effects estimate, not how inflation responds to slack. Independent evidence for the paper’s central premise comes from [Alvarez-Blaser et al. \(2025\)](#), who document with barcode-level data that disaggregated inflation loads heterogeneously on latent common factors, with differential transmission first-order at the product-category level; they measure how much of inflation’s variance this granular structure explains, whereas we ask what the same structure does to the identification of the Phillips curve slope.

On the *method*, it builds on panel models with interactive fixed effects ([Bai, 2009](#)). It speaks most directly to the literature on recovering micro-local¹⁰ elasticities from panels, whose standard regional design is codified in [Chodorow-Reich \(2020\)](#): the balancing and shift-share designs of [Goldsmith-Pinkham et al. \(2020\)](#), [Borusyak et al. \(2022\)](#), [Arkhangelsky and Korovkin \(2024\)](#) and [Manuel \(2026\)](#), the external-instrument and general-equilibrium-decomposition designs of [Donaldson \(2025\)](#) and [McKay and Wolf \(2023\)](#), the regional-SVAR approach of [Sarto \(2025\)](#), which also exploits a factor structure to estimate the aggregate shock from within the panel, and work on inference under cross-sectional dependence ([Almuzara and Sancibrián, 2024](#)). Our contribution to this group is twofold: we identify an inference problem that is distinct from the point estimate bias these methods target and is not repaired by recentering the estimate, and we establish that interactive fixed effects resolve the bias and inference failures using only the granularity of the panel, requiring neither an observed shock nor observed exposure measures, nor the exclusion restrictions on a policy block that internal shock construction demands. For the synchronized slope, Section 7.4 adds an exclusion restriction, whereby innovations to the regressor’s common component are orthogonal to shocks entering inflation directly, allowing a causal interpretation of that response.

A detailed placement of the paper within these literatures – including the formal comparison of identification routes and the connection to the granular-instrumental-variables approach – is given in Online Appendix [OA.1.1](#).

The main alternative for a multifactor error structure is the common correlated effects (CCE) estimator of [Pesaran \(2006\)](#), which proxies the unobserved factors with cross-sectional averages of the outcome and the regressor. Section 7.6, with Experiment G14 in Appendix C,

⁹Some studies use the survey measures to approximate vacancies or analyze aggregate only wage-Phillips curves ([Kiss et al., 2022](#); [Bonam et al., 2021](#); [Bonthuis et al., 2016](#)). Crucially, our approach exploits sub-sector-country granularity to link labor shortages to inflation (producer prices) dynamics.

¹⁰The term of [Sarto \(2025\)](#), who distinguishes the micro-local from the general-equilibrium-inclusive micro-global elasticity; “portable,” in the language of [Nakamura and Steinsson \(2018\)](#).

compares it – together with the mean-group estimator of [Pesaran and Smith \(1995\)](#) – with IFE. CCE fails when the loadings average to zero across units, when the panel has as many factors as averages or when the averages capture the wrong aggregate – and none of these can be tested, whereas an omitted factor under IFE shows up in the rank criteria. It also recovers the micro-local slope α_x only, with no counterpart for α_c . Only where the averages track the factors does it perform nearly as well as IFE.

On *inference*, the cross-sectional roots of our result lie in [Andrews \(2005\)](#): with common shocks, the least-squares limit is random conditional on them. His result concerns estimation in a single cross-section, while ours arises from pooling – each added period re-estimates the same realization of exposures, so the reported standard error falls while the true uncertainty does not. Clustering is also the solution for Moulton’s problem ([Moulton, 1986, 1990](#)), but the two are different. In Moulton’s case, errors are correlated within a group, so the sample carries less information than it appears to – yet more data still help. In the context analyzed here they do not: each sector’s exposure to the common cycle is a fixed characteristic that every quarter re-measures, so clustering by unit reveals an interval that a longer panel never narrows (Appendix B.4). [Bertrand et al. \(2004\)](#) also find that more periods make inference worse, but through serial dependence of the errors; our channel is cross-sectional dispersion of the loadings, so the two are complements rather than substitutes.

3 The Setting and Three Failures of Time Fixed Effects

The environment. Let $i = 1, \dots, N$ index cross-sectional units and $t = 1, \dots, T$ index time periods. The outcome and the regressor of interest are each driven by a common factor and by idiosyncratic variation:

$$y_{it} = \alpha x_{it} + \lambda_i F_t + \gamma_i + \varepsilon_{it}, \quad (1)$$

$$x_{it} = \mu_i F_t + \xi_{it}. \quad (2)$$

Here, F_t is a scalar common factor capturing the aggregate forces that move all units together; λ_i is the sensitivity of unit i ’s outcome to that factor and μ_i the sensitivity of its regressor; γ_i is a unit fixed effect; and ε_{it} and ξ_{it} are mean-zero idiosyncratic disturbances. The two loadings λ_i and μ_i are distinct objects: the same aggregate shock may pass strongly into a unit’s outcome but weakly into its regressor, or vice versa. This means that a common energy shock may pass almost one-for-one into the producer prices of German chemicals, a capital-intensive industry whose hiring needs change little – a large λ_i next to a small μ_i – while Portuguese apparel may show only a small price response to the same shock even as

its hiring difficulties move sharply – a small λ_i alongside a large μ_i . The bias of the time fixed effects estimator depends on how these two sensitivities co-vary across units.

The setup is general: nothing requires the outcome and the regressor to load on the *same* aggregate or on a single one. Writing both equations with one common factor F_t is a presentational choice we make for expositional clarity – it lets each failure be isolated analytically – and the results below do not depend on it. Where the distinction becomes relevant, we write F_t^y for the outcome’s aggregate and F_t^x for the regressor’s.

Loadings and dynamics. The unit-specific loadings (λ_i, μ_i) are drawn i.i.d. across units from a bivariate distribution with means $(\bar{\lambda}, \bar{\mu})$, standard deviations $(\sigma_\lambda, \sigma_\mu)$ and correlation $\rho_{\lambda\mu}$. The correlation $\rho_{\lambda\mu}$ is a central parameter of the analysis: it measures the degree to which units whose regressor responds strongly to the aggregate also have outcomes that respond strongly to it. Throughout, we take the symmetric case $\sigma_\lambda = \sigma_\mu \equiv \sigma$ as the baseline, so that a single parameter σ indexes the overall degree of loading heterogeneity and $\rho_{\lambda\mu}$ its directional alignment; the asymmetric case adds no qualitative insight and is carried implicitly.

Bridging the general notation with the Phillips curve application, $\rho_{\lambda\mu} > 0$ says that the units whose measured slack co-moves most strongly with the aggregate cycle are also the units whose inflation responds most strongly to it¹¹. A New Keynesian model delivers exactly this pattern if energy enters production with different intensities across units – an energy-hungry sectors’ marginal costs and hiring both answer the common energy price – or if inflation expectations are forward-looking and formed with different sensitivity to aggregate conditions across units.

The common factor follows a stationary first-order autoregression, $F_t = \rho_F F_{t-1} + u_t$ with $|\rho_F| < 1$, as does the idiosyncratic component of the regressor, $\xi_{it} = \rho_\xi \xi_{i,t-1} + v_{it}$; the innovations u_t and v_{it} are mean-zero and mutually independent. The outcome disturbance ε_{it} is mean-zero, independent of the loadings and of the factor path $\{F_t\}$, and orthogonal to the idiosyncratic component of the regressor at all leads and lags, $\mathbb{E}[\xi_{is} \varepsilon_{it}] = 0$ for all s, t – the idiosyncratic-exogeneity condition. This holds, for instance, whenever ε_{it} is conditionally mean-zero given the loadings, the factor path and the entire history of ξ_i ; it is the panel analogue of regressor exogeneity with respect to the variation that survives the within transformation. Practically this means that, conditional on the common cycle and unit exposure to it, the remaining unit-specific variation in slack is unrelated to the remaining inflation

¹¹Conversely, if $\rho_{\lambda\mu} < 0$, the units whose measured slack is most procyclical are those whose inflation responds least strongly to the common cycle.

disturbance.

The estimands. The decomposition of the regressor into idiosyncratic and synchronized variation gives rise to two economically distinct responses. The first is the micro-local slope, α_x : the response of the outcome to a unit-specific movement in the regressor, holding the common state fixed. It is the effect a researcher would recover using only the idiosyncratic variation in x_{it} . In the application, it asks how much inflation in one unit changes when tightness in that pair rises, while tightness elsewhere and the common forces driving inflation are held fixed.

The second is the synchronized slope, α_c : the response to a movement in tightness shared across units. This is a different experiment. A local tightening leaves aggregate conditions essentially unchanged; a synchronized tightening moves many labor markets at once and may therefore operate through persistence, aggregate expectations, wage norms or other general-equilibrium forces. Inflation need not respond to the two movements in the same way. In the baseline model (1), this distinction is suppressed by imposing $\alpha_c = \alpha_x = \alpha$. The extension in (4) relaxes that restriction.

Once α_c and α_x are allowed to differ, a single-slope specification generally identifies a combination of the two coefficients. Even when heterogeneous exposure creates no omitted variable bias, Lemma 2 shows that the coefficient is a weighted average of the two responses, with the weight determined by how much synchronized variation the estimator retains. Heterogeneous exposure introduces a separate distortion by correlating the remaining common component of the regressor with the outcome disturbance. The conventional time fixed effects coefficient can therefore differ from the micro-local slope both because it combines two distinct responses and because heterogeneous exposure biases that blend.

The two slopes also differ in their identification. Under the idiosyncratic-exogeneity condition, the defactored variation identifies α_x . The synchronized component identifies α_c only as a projection unless aggregate tightness is instrumented; under the exclusion restriction introduced below, the instrument gives α_c a causal interpretation. On that interpretation, $\alpha_c - \alpha_x$ captures the aggregate response that cross-sectional designs difference out – the object associated with the “missing-intercept” problem¹².

¹²See [Wolf \(2023\)](#) for a formal treatment and [Moll and Hanney \(2025\)](#) for an accessible introduction.

3.1 Failure I: Point Estimate Bias

Time dummies remove, period by period, the *cross-sectional average* of the common component – the term $\bar{\mu}F_t$ in the regressor and $\bar{\lambda}F_t$ in the outcome – but not its cross-unit transmission. The unit effects remove each series’ own time mean. Once unit and time effects are netted out, the regressor continues to include $(\mu_i - \bar{\mu})\tilde{F}_t$, whereas the error term includes $(\lambda_i - \bar{\lambda})\tilde{F}_t$. Take an increase in euro area energy prices as an example. Time dummies capture the average inflation reaction, but prices can rise more in energy-intensive industries. If hiring difficulties also intensify more in those industries, the residual variation in slack becomes correlated with the error term. As a result, the regression assigns to slack some of the inflation response that is actually driven by exposure to the energy shock.

Lemma 1 (Time FE Bias). *Under the DGP (1)–(2), with (λ_i, μ_i) i.i.d. across units with variance σ^2 , and correlation $\rho_{\lambda\mu}$; F_t strictly stationary with $\text{Var}(F_t) > 0$; ξ_{it} and ε_{it} mean-zero, independent of all loadings and of $\{F_t\}$; $\mathbb{E}[\xi_{is}\varepsilon_{it}] = 0$ for all s, t (idiosyncratic exogeneity); and $\text{Var}(\tilde{x}_{it}) > 0$:*

$$\text{plim}_{N,T \rightarrow \infty}(\hat{\alpha}_{\text{TimeFE}} - \alpha) = \frac{\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) \cdot \text{Var}(\tilde{F}_t)}{\text{Var}(\tilde{x}_{it})}, \quad (3)$$

where tildes denote residuals after projecting out unit and time means. Under i.i.d. loadings, $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = \rho_{\lambda\mu}\sigma^2(1 - N^{-1})$, so under idiosyncratic exogeneity the bias equals zero if and only if $\rho_{\lambda\mu}\sigma^2 = 0$.

Proof. See Appendix B.1 □

Two features of (3) organize what follows. The bias inherits the sign of the loading correlation $\rho_{\lambda\mu}$ and scales with the degree of loading heterogeneity σ^2 : heterogeneity sets the magnitude, the correlation sets the direction. A panel with substantial dispersion in how units load on the aggregate, but with that dispersion uncorrelated across the two equations, is unbiased; a panel with even modest dispersion that is correlated across equations is not.

This dispersion (σ^2) has a familiar New Keynesian reading. Suppose the frequency of price adjustment is higher in German chemicals – where prices track input costs that move daily – than in Portuguese apparel, which reprices by season and catalogue. Then the same euro area-wide cost shock – an energy price jump, a common shift in expectations – passes into the first industry’s inflation faster and to a greater degree than into the second’s: the two respond differently to the *same* aggregate movement. This corresponds to heterogeneity in units’ loadings on the aggregate and follows directly from allowing standard Calvo parameters to

vary across units. The correlation ($\rho_{\lambda\mu}$) that drives the bias in Lemma 1 is equally familiar. If energy enters the production function with different intensities across units, what the dummies leave behind is each sector’s above- or below-average sensitivity to the common energy shock – and since the units whose marginal costs, and hence inflation, overreact to that shock may also be the sectors whose slack overreacts to it, that shared overreaction is exactly the surviving correlation that biases the slope. Heterogeneous responses to inflation expectations can generate a similar bias: if they are forward-looking and formed with different sensitivity to aggregate conditions, the sectors that reprice most on common news may also be the sectors whose hiring difficulties respond most. Either channel alone is enough to bias the slope; the application will measure, rather than assume, how strongly the two sensitivities align.

Lemma 1 is stated with a single common factor for exposition only. A comprehensive two-aggregate treatment can be found in Corollary 1, in Appendix B.5.

3.2 Failure II: Estimand Conflation with Two Aggregates

The second identification problem arises when the regressor carries a common component of its own, and it arises even when Lemma 1 delivers no bias at all. In our Phillips curve setting this will be most relevant: in the empirical application, the common components of slack and inflation are far from perfectly aligned, with an estimated correlation of only 0.3.

Write F_t^y for the factor in the outcome equation and F_t^x for the factor driving the regressor, so that each unit’s regressor is the sum of a *synchronized* component $\mu_i F_t^x$ – the economy-wide movement scaled by the unit’s exposure – and an *idiosyncratic* component ξ_{it} . In the Phillips curve reading the two aggregates are distinct objects: F_t^y collects the forces time dummies were introduced for – shared inflation expectations, energy and other cost-push shocks, the common stance of policy – while F_t^x is the common component of measured tightness, the hiring cycle the euro area shares.

Once the two components are separated, there is no reason for inflation to respond equally to them. An idiosyncratic tightening affects one sector in isolation; a synchronized tightening affects many sectors at once and can therefore operate through aggregate capacity constraints, inflation expectations and wage-setting. The two slopes may therefore differ. When the two responses differ, any estimator that imposes a single slope necessarily combines them. The resulting coefficient is a weighted average of α_x and α_c , with the weights determined by how much idiosyncratic and synchronized variation remains after the estimator removes common components. Different estimators therefore target different combinations of the two

slopes. This is distinct from the bias in Lemma 1: the problem is not that a common slope is estimated incorrectly, but that no single common slope exists.

The regressor is now decomposed as $x_{it} = \mu_i F_t^x + \xi_{it}$, with the regressor's own aggregate in place of F_t , and the outcome is allowed to respond to its two components with distinct coefficients,

$$y_{it} = \alpha_c \mu_i F_t^x + \alpha_x \xi_{it} + \lambda_i F_t^y + \gamma_i + \varepsilon_{it}. \quad (4)$$

Setting $\alpha_c = \alpha_x$ and $F_t^x = F_t^y$ returns the baseline model; Appendix B.5 states the two-factor environment in more detail. The two coefficients distinguish the source of an increase in a sector's tightness. α_x measures the inflation response to a one-unit increase confined to that sector. α_c measures the response to the same increase when it reflects a rise in the common tightness factor F_t^x , with other sectors affected according to their exposures. Both responses hold the common inflation forces F_t^y fixed. A time fixed effects regression uses tightness relative to the cross-sectional average, which still contains both sources of variation. As the lemma below shows, its coefficient therefore combines the two responses.

Consider an estimator e that absorbs a set of effects and then fits a single linear slope. Write $\tilde{x}_{it}^{(e)}$ for the regressor variation that survives its absorption. We consider two absorption schemes. *Time fixed effects* (TFE) absorb unit and period means, so $\tilde{x}_{it}^{(\text{TFE})}$ is the two-way residual of Lemma 1. *Factor absorption* (FA) absorbs unit means and, unit by unit, the projection of the regressor on the outcome's aggregate F_t^y , so $\tilde{x}_{it}^{(\text{FA})}$ is the residual of x_{it} from a unit-specific regression on a constant and F_t^y . For any scheme e , define the *synchronized-component weight*

$$w_e \equiv \frac{\text{Cov}\left(\tilde{x}_{it}^{(e)}, \mu_i F_t^x\right)}{\text{Var}\left(\tilde{x}_{it}^{(e)}\right)}, \quad (5)$$

the share of the surviving variation that is synchronized movement. Say that e satisfies the *zero-bias condition* if the residual factor structure it leaves behind is orthogonal to $\tilde{x}_{it}^{(e)}$, so that the omitted variable term of Lemma 1 vanishes. The following result formalizes the conflation¹³.

Lemma 2 (Estimand conflation across components of the regressor). *Let (4) hold with (F_t^y, F_t^x) jointly strictly stationary with absolutely summable cross-autocovariances, ξ_{it} and ε_{it} mean-zero and independent of the loadings and of both factor paths. The remaining*

¹³Its form will look somewhat familiar to some readers: pooled linear estimators recover design-weighted averages of heterogeneous responses, the lesson of the decomposition literature on two-way fixed effects (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeulle, 2020; Słoczyński, 2022).

conditions of Lemma 1 are in force. Then the micro-local slope defined above, conditional on both factors, equals α_x , and for any estimator e satisfying the zero-bias condition with $\text{Var}(\tilde{x}_{it}^{(e)}) > 0$,

$$\text{plim } \hat{\alpha}_e = w_e \alpha_c + (1 - w_e) \alpha_x, \quad (6)$$

so that

$$\text{plim } \hat{\alpha}_{\text{FA}} - \text{plim } \hat{\alpha}_{\text{TFE}} = (\alpha_c - \alpha_x) (w_{\text{FA}} - w_{\text{TFE}}). \quad (7)$$

The estimator targets α_x only if $w_e = 0$ or $\alpha_c = \alpha_x$.

Proof. See Appendix B.2. □

A panel Phillips curve can therefore combine two distinct relationships. A sector may experience labor shortages specific to its own market or as part of a broader tightening across the euro area. Inflation need not respond equally to the two. Sector-specific shortages may ease as workers move across sectors, whereas widespread shortages limit this adjustment and may generate more persistent cost pressures. The corresponding slopes are α_x and α_c . A regression imposing a single slope combines them according to the share of synchronized variation remaining after the fixed effects are removed. This conflation can arise even when $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = 0$, so it is distinct from the omitted variable bias in Lemma 1: a single coefficient combines the two responses even in the absence of confounding.

There are several reasons why the two responses may differ. Synchronized tightness is typically more persistent than a local shortage and may interact with aggregate expectations, wage-setting, and other economy-wide forces in ways that idiosyncratic tightness does not. We return to these interpretations, and to the identification of the synchronized response, in Section 7.4.

Remark 1 (Signal deletion: what time demeaning removes). Period dummies remove the cross-sectional *average* of the regressor's common component, $\bar{\mu} F_t^x$, but keep each unit's deviation from it, $(\mu_i - \bar{\mu}) F_t^x$. Factor absorption, which projects on the *outcome's* aggregate, retains the regressor's synchronized variation whenever the two aggregates differ. Hence $w_{\text{TFE}} < w_{\text{FA}}$, when the two aggregates are weakly correlated, and the deletion changes *what* is estimated, not merely how precisely; yet neither weight is zero, since the dummies delete only the average. Appendix B.5 develops the mechanism.

Remark 2 (The two-step estimator). Lemma 2, together with the defactoring corollary of Appendix B.5 (Corollary 1), licenses a two-step procedure, and the two steps are not symmetric.

Step one – defactor and recover α_x . We first separate each unit’s tightness into its synchronized and idiosyncratic components. Applying principal components to the within-transformed regressor panel, with the rank selected by the same criteria used for the outcome, yields an estimate of the common factor F_t^x and of each unit’s exposure μ_i . The fitted component $\hat{\mu}_i \hat{F}_t^x$ captures the part of tightness driven by the common cycle; subtracting it from x_{it} leaves the unit-specific component.

In the application, this amounts to removing the euro area hiring cycle from each unit. One that typically moves strongly with that cycle has more removed; one that moves little has less removed. What remains is therefore each unit’s tightness relative to the common cycle. By construction, this defactored regressor contains no synchronized component, so its weight in (5) is zero. Corollary 1(ii) then implies that regressing inflation on this residualized variation recovers α_x , the micro-local response of inflation to a unit’s own tightness.

Implementation is straightforward: the common component is estimated from the regressor panel itself, and recovering α_x requires no external instrument¹⁴.

Step two – identify the synchronized response. Once the common component of tightness has been recovered, the remaining question is how inflation responds to it. Although $\mu_i F_t^x$ varies across units through the loading μ_i , its underlying source of variation is the common time series F_t^x . The cross-section therefore helps estimate exposure to aggregate tightness, but it does not create independent aggregate shocks. Identification of the synchronized response remains, in this sense, a time-series problem.

This distinction matters for interpretation. Regressing inflation on $\mu_i F_t^x$ yields a projection coefficient α_c , but that coefficient need not be causal: the same demand, cost or policy forces that move aggregate tightness may also affect inflation directly (McLeay and Tenreyro, 2020). A causal interpretation therefore requires an instrument z_t that shifts F_t^x while remaining orthogonal to the aggregate forces entering inflation¹⁵. The panel remains useful though, as heterogeneous loadings turn a single aggregate innovation into differential first-stage variation across units¹⁶. But the identifying shock itself remains aggregate.

Finally, α_c should not be confused with a policy counterfactual. It is the response to synchro-

¹⁴Because the common component is estimated, defactoring is imperfect in finite samples and may leave some synchronized variation in the residualized regressor. This disappears asymptotically; in the application, the bootstrap accounts for the estimation error by re-estimating the factor structure in each replication.

¹⁵Relevance requires $\bar{\mu} \text{Cov}(\tilde{z}_t, \tilde{F}_t^x) \neq 0$, while exclusion requires $\text{Cov}(\tilde{z}_t, \tilde{F}_t^y) = 0$. Granularity, Assumption 1, delivers orthogonality to idiosyncratic disturbances. With several regressor factors, one instrument is required for each aggregate component.

¹⁶The same logic determines inference. Because the synchronized response is identified from aggregate time-series variation, its inference is governed by T , not NT ; Appendix C studies the implications.

nized tightness conditional on the outcome’s own common factor, not the total response of aggregate inflation to a policy intervention (Wolf, 2023; Moll and Hanney, 2025). Section 7.4 therefore reports two objects: the projection coefficient α_c , which may still reflect aggregate forces correlated with synchronized tightness, and its causal counterpart obtained with the internal instrument of Sarto (2025) (Appendix D.6).

3.3 Failure III: Inference Collapse

The inference problem remains even when the point estimate is unbiased. Set $\rho_{\lambda\mu} = 0$, so that Lemma 1 delivers no asymptotic bias, and let the outcome respond to the synchronized and idiosyncratic components with the same slope. After time demeaning, the residual still contains $(\lambda_i - \bar{\lambda})\tilde{F}_t$: observations from different units therefore remain linked by the same realization of the common factor, scaled by their unit-specific exposures.

Lemma 3 (Variance floor). *Under the conditions of Lemma 1, with finite fourth moments and the weak-dependence conditions of Appendix B.3,*

$$\sqrt{N}(\hat{\alpha}_{\text{TimeFE}} - \alpha) \xrightarrow{d} \mathcal{N}(0, V), \quad V = \frac{\kappa[\text{Var}(\tilde{F}_t)]^2}{[\text{Var}(\tilde{x}_{it})]^2}, \quad (8)$$

where $\kappa = \text{Var}(\tilde{\lambda}_i\tilde{\mu}_i)$ is assumed strictly positive; under independent loadings $\kappa = \sigma^4$.

The source of the $\sqrt{N}$ rate is transparent from the leading term in the estimation error:

$$\underbrace{\left(\frac{1}{N} \sum_i \tilde{\lambda}_i \tilde{\mu}_i\right)}_{\substack{\text{one number per unit:} \\ \text{its two exposures, multiplied}}} \cdot \underbrace{\left(\frac{1}{T} \sum_t \tilde{F}_t^2\right)}_{\substack{\text{the cycle's variance} \\ \text{over the sample}}} + \underbrace{\text{remainder}}_{\substack{\text{averages away over} \\ \text{all } NT \text{ observations}}}.$$

As T grows, the second bracket converges to $\text{Var}(\tilde{F}_t)$ rather than to zero, while the first is unaffected because the unit-specific exposures are fixed across time. A longer panel therefore observes the same cross-section of exposure products more precisely but does not generate new draws of them. Only a wider cross-section reduces this source of uncertainty, so the relevant sampling error is of order $N^{-1/2}$ and the estimator converges at rate $\sqrt{N}$.

Conventional standard errors nevertheless shrink as if all NT observations supplied independent information and therefore miss this floor. Clustering by unit captures this uncertainty by allowing observations from the same unit to remain correlated over time. Each unit’s loadings are fixed across periods, so extending the panel does not provide new draws of those exposures. Unit clustering accounts for their dispersion across units

and therefore captures the variance floor. Corrections based on period-level scores instead measure variation across dates and miss this component. Clustering thus restores valid inference under the conditions of Lemma 3, but does not remove the floor or correct the bias when loadings are correlated. Appendix B.3 provides the full derivation and coverage result, and Appendix B.4 compares alternative variance estimators.

In the applications that motivate this paper, the three problems are likely to arise together. A panel regression of unit inflation on unit slack with period fixed effects can bias the slope when the same characteristics that shape a unit’s exposure to aggregate slack also shape its inflation response – for example, energy intensity, price-setting frequency or sensitivity to inflation expectations. Even absent that bias, the coefficient can blend the responses to idiosyncratic and synchronized tightness. And conventional confidence intervals can become increasingly precise with the length of the panel while remaining too narrow. Each problem has a different source, but the two-step extension to interactive fixed effects address all three.

4 IFE: Assumptions and Main Result

The three distortions share a common source: time fixed effects represent common forces by their cross-sectional means, even when units are exposed heterogeneously to these forces. Our extension to interactive fixed effects (Bai, 2009) separates the regressor into synchronized and idiosyncratic components, while jointly estimating the slope, common factors and unit-specific loadings.

In the notation of Section 3.2, standard IFE converges to the factor-absorption problem: the identifying variation is $\tilde{x}_{it}^{(\text{FA})}$, and the corresponding weight w_{FA} in Lemma 2 is henceforth denoted w_{IFE} . Defactoring the regressor removes its synchronized component altogether, setting this weight to zero and leaving the variation that identifies α_x . The following five assumptions are sufficient for this procedure to recover the micro-local slope.

Four regularity conditions, standard in the factor-model literature and stated formally as Assumptions 2–5 in Appendix A, are maintained throughout¹⁷: loadings vary genuinely (no two sectors answer the common cycle identically), the factor is strong enough to be detected (the common inflation cycle is pervasive rather than confined to a handful of sectors), it is not collinear with the unit effects (the cycle moves over time instead of being a fixed sectoral trait) and the idiosyncratic errors do not themselves form a second factor (whatever co-movement

¹⁷Those assumptions are borrowed from Bai (2009).

survives the estimated cycle is local). The next assumption is this paper's addition. It permits feedback from individual units to the aggregate while preserving identification from cross-sectional variation, and it is directly testable by checking that no single unit carries too much weight in the panel.

Assumption 1 (Granularity). *Let $s_i \geq 0$ denote unit i 's share of the weighted aggregate $\bar{X}_t^w \equiv \sum_i s_i x_{it}$, with $\sum_i s_i = 1$, and define the Herfindahl–Hirschman index*

$$\text{HHI} \equiv \sum_{i=1}^N s_i^2. \quad (9)$$

The panel satisfies the granularity assumption if $\max_i s_i = O(1/N)$, and hence $N \cdot \text{HHI} = O(1)$. Under this assumption, for any aggregate M_t responding to $\bar{X}_t^w$ with finite feedback coefficient ψ ,

$$|\text{Cov}(\xi_{it}, M_t)| \leq |\psi| s_i \sigma_\xi^2 \rightarrow 0. \quad (10)$$

For the euro area panel, granularity requires that no single unit – the German motor vehicles sector, for example – carries enough economic weight for its own hiring difficulties to move aggregate euro area slack. Each unit can therefore treat the aggregate as given. The condition concerns economic weights, not factor loadings, and can be checked directly in the data. In the application it will be done using the Commission's output weights (Section 7).

The granularity condition also has a useful finite-sample interpretation: it bounds the bias generated by general-equilibrium feedback in terms of quantities that can be observed or disciplined in the data.

Remark 3 (Operational threshold for granularity). For fixed N , suppose general-equilibrium feedback of strength ψ generates an omitted second factor with outcome loading θ_i . The resulting bias in $\hat{\alpha}_{\text{IFE}}$ is bounded by

$$|\text{plim}(\hat{\alpha}_{\text{IFE}} - \alpha)| \lesssim |\psi| \max_i s_i \bar{\theta}, \quad \bar{\theta} \equiv N^{-1} \sum_i |\theta_i|. \quad (11)$$

The result follows by applying Lemma 1 to the component of the feedback that remains after rank-one factor absorption. Feedback operating through F_t is absorbed into the estimated loadings (Appendix B.6); the remaining factor is the share-weighted average of idiosyncratic shocks, $G_t \equiv \bar{\xi}_t^w$. The outcome loads on this factor through $\psi\theta_i$, while the regressor exposure

is s_i/HHI . Substituting these terms into the bias formula gives

$$\text{plim}(\hat{\alpha}_{\text{IFE}} - \alpha) = \psi \text{Cov}(s_i, \theta_i),$$

which yields (11). Thus the residual bias depends directly on the feedback strength and the concentration of economic weights: as the largest unit share becomes small, so does the scope for feedback to contaminate the slope. Using the post-projection variance σ_ξ^2 in the denominator makes the bound conservative, since any additional variation left in the residualized regressor can only weaken it.

The granularity assumption is operationally satisfied whenever

$$|\psi| \cdot \max_i s_i \ll \frac{|\alpha|}{\theta}. \quad (12)$$

The threshold has two independent levers: strong GE feedback is innocuous if the panel is sufficiently disaggregated and even a mildly dominant unit causes problems only if the feedback is appreciable. Throughout, the shares s_i are *economic* weights, and the diagnostics of Appendix D.7 are computed with them¹⁸. We provide some intuition below.

To summarize, the first four assumptions are standard in the factor-model literature and provide the standard conditions for consistent estimation of the common structure. Assumption 1 is the additional requirement introduced here. It allows unit-level shocks to feed back into the aggregate while ensuring that, under the maintained factor rank, no single unit contributes enough to generate an additional common component. The relevant quantity is $|\psi| \max_i s_i$ (Remark 3), so the condition can be assessed directly from the concentration of unit weights.

Granularity should therefore be viewed as a sufficient condition for the baseline rank specification, rather than as a requirement that general-equilibrium feedback be absent. Monte Carlo simulations (Table A.9) shows that moderate departures from the condition generate little distortion. When concentration becomes large enough for feedback to matter, the feedback itself increasingly appears as an additional common factor. Allowing the factor rank to increase then removes the distortion. Thus the rank criteria provide a useful diagnostic when granularity is imperfect, but they do not replace the assumption underlying the baseline result.

¹⁸When weights are not readily available, the equal-weighted benchmark $s_i = N^{-1}$ (so $\text{HHI} = N^{-1}$) reduces the bound to $|\psi|\bar{\theta}/N$, which vanishes at the parametric rate regardless of feedback strength. Equal weights may assume away concentration, however: per Gabaix (2011), a large N does not by itself imply a small weighted Herfindahl, so the equal-weighted number should be read as a lower bound.

In the disaggregated panels considered here, concentration is low by construction, making granularity particularly plausible precisely in settings where heterogeneous exposure to common shocks is most pronounced.

Proposition 1 (IFE Identification). *Therefore, under Assumption 1 and Assumptions 2–5, with $N, T \rightarrow \infty$ such that $T/N \rightarrow \varrho \in (0, \infty)$, the IFE estimator $\hat{\alpha}_{IFE}$ is:*

- (i) *consistent: $\hat{\alpha}_{IFE} \xrightarrow{p} \alpha$;*
- (ii) *$\sqrt{NT}$ -asymptotically normal¹⁹;*
- (iii) *strictly preferred to $\hat{\alpha}_{TimeFE}$, in the sense that $|\text{plim}(\hat{\alpha}_{TimeFE} - \alpha)| > 0$ whenever $\rho_{\lambda\mu} \neq 0$, while the nominal 95% coverage of $\hat{\alpha}_{TimeFE}$ falls below 0.95 whenever $\sigma^2 > 0$.*

Proof. See Appendix B.6 □

Implementation details connecting Proposition 1 to estimation – the factor normalization, the treatment of the additive effects, the residualized-regressor rank and generated-factor standard errors – are collected in Remark 7 in Appendix B.6.

Selecting the factor rank. The assumptions above take the number of factors as known; in practice it must be estimated. We select the rank r from the residuals of the baseline fixed effects regression using three complementary criteria, all building on the information-criteria framework of Bai and Ng (2002): the eigenvalue-ratio and growth-ratio criteria of Ahn and Horenstein (2013); the criterion of Onatski (2010) and the diagnostic of Gagliardini et al. (2019). Using several criteria guards against the known tendency of any single one to over- or under-select in finite samples. Details on the construction and implementation of these rank-selection criteria are relegated to Online Appendix OA.1.3.

The following remark summarizes the practical steps for implementing the estimator and readers may return to it when considering its empirical application.

Remark 4 (Implementation checklist). Consider a panel outcome y_{it} and regressor x_{it} . Time fixed effects (time FE) absorb unit and period means. Interactive fixed effects (IFE) instead model common variation with estimated factors and unit-specific loadings. If x_{it} has its own common component, write it as the sum of synchronized variation,

¹⁹With variance as characterised in Bai (2009, Theorem 3).

$\mu_i F_t^x$, and an idiosyncratic remainder, ξ_{it} . The response to the latter is the micro-local slope α_x ; the response to the former is α_c .

1. *Diagnose common structure in the regressor.* Apply principal component and factor-rank criteria to the within-transformed regressor panel (à la [Bai and Ng, 2002](#)). If x_{it} contains an economically meaningful common component, retain the selected factor F_t^x and its rank for the decomposition below.
2. *Diagnose common structure in the outcome and estimate IFE.* Estimate the time-FE specification and apply the same rank-selection criteria to its residuals. If no meaningful common component is present, no factor correction of the outcome is needed. Otherwise, estimate the IFE specification of [Bai \(2009\)](#) at the selected rank r , alongside the conventional time-FE regression. A large difference makes a case for modelling heterogeneous exposure to common forces.
3. *Check granularity.* Using economic weights s_i , compute $\max_i s_i$ and $\text{HHI} = \sum_i s_i^2$. Then re-estimate after dropping the largest units and compare the estimated factors with the full-sample factor. Small concentration and a stable factor support Assumption 1; instability indicates that a few units may be feeding materially into the aggregate²⁰.
4. *Determine whether the regressor has a distinct aggregate.* Compare the regressor factor F_t^x from Step 1 with the estimated outcome factor from Step 2. Whether the two should be treated as distinct depends jointly on their correlation and on the strength of the common component in the regressor. Experiment G15 (Table [A.13](#)) provides the corresponding decision thresholds.
5. *Decompose the regressor.* Estimate each unit's loading μ_i on F_t^x and split x_{it} into the fitted synchronized component $\hat{\mu}_i \hat{F}_t^x$ and the idiosyncratic remainder, $x_{it} - \hat{\mu}_i \hat{F}_t^x$.
6. *Recover the two slopes.* Regress the outcome on both components jointly, retaining the IFE structure of Step 2 so that $\lambda_i F_t^y$ is controlled for. α_x is the coefficient of the idiosyncratic component and α_c of the synchronized one. Because the components are estimated, obtain standard errors from a bootstrap that re-estimates the factor structure in every replication.
7. *Instrument α_c for a causal interpretation.* Instrument the synchronized component

with an innovation to the regressor’s common factor (internal) or an externally identified aggregate shock (external), interacted with the estimated loadings. The identifying variation is a time series, so inference should be at the time level, allowing for serial dependence.

8. *Report the four objects separately.* Report the conventional time-FE coefficient with standard errors clustered by unit, the IFE coefficient at the selected rank, and, when the regressor has a distinct common component, the two component slopes α_x and α_c , distinguishing the projection estimate of α_c from its IV counterpart.

A sample routine implementing these steps accompanies the paper.

Which route and when. Interactive fixed effects are one among other designs for a panel with common shocks and the checklist’s own diagnostics decide when they apply: Steps 1–2 ask whether a factor structure is present at all, and Step 3 whether the panel is granular. Table [OA.2](#) provides further details on the choice between approaches and when each is appropriate.

5 Monte Carlo Evidence

We validate and extend the main analytical results of Sections [3–4](#) in a sequence of Monte Carlo simulations. The data are generated from the two-factor-loading model of Section [3](#), and the true slope is fixed at $\alpha_x = 1.0$ in every experiment, so that bias is read directly as the gap between an estimator’s mean and unity, and coverage is the share of nominal 95% intervals that contain the actual coefficient²¹. Every replication draws a fresh set of loadings $\{\lambda_i, \mu_i\}$ alongside the factor path and the idiosyncratic shocks, so these simulations implement the unconditional experiment (i) of Remark [5](#). The reported coverage rates refer to repeated draws of both loadings and shocks; the remark discusses their interpretation for a single realized panel. We report one exhibit for each main result: the bias and inference failures and their behaviour as the panel lengthens (Experiment G1); the point estimate bias tracking the loading correlation, and its resolution (Experiment G2); the conflation that both pooled estimators inherit once the responses to the two components of the regressor differ

²⁰When economic weights are unavailable, the equal-weight HHI and a leave-one-out exercise provide useful diagnostics.

²¹We remind the reader that coverage denotes the fraction of Monte Carlo replications in which the nominal 95% confidence interval $\hat{\alpha} \pm 1.96 \widehat{\text{se}}(\hat{\alpha})$ contains the true slope $\alpha = 1$.

(Experiment G3) and its removal by the two-step procedure (Experiment G4); the recovery of the synchronized slope by the instrumented second step (Experiment G5); and the cost of measurement error in the regressor – the boundary that motivates the slack measure of Section 6 (Experiment G6). Additional simulations are collected in Appendix C and summarized at the end of this section: general-equilibrium feedback and the role of the factor rank (Experiment G7); the granularity dividend (Experiment G8); the inference side of the instrumented second step – exact identification of α_c , without a reliable interval at realistic T (Experiment G5, continued); the evidence that no variance correction restores Time FE coverage (Experiment G9); and a direct comparison with the mean-group and common correlated effects estimators of Pesaran and Smith (1995) and Pesaran (2006) (Experiment G14), among others.

5.1 Two-Way Panel Design

Design. The data are generated from $y_{it} = \alpha x_{it} + \lambda_i F_t + \gamma_i + \varepsilon_{it}$ and $x_{it} = \mu_i F_t + \xi_{it}$, with $N = 250$ cross-sectional units, $\alpha = 1.0$, $\rho_F = 0.80$, and $\rho_\xi = 0.80$. The loadings (λ_i, μ_i) are drawn from a bivariate distribution with common heterogeneity σ and correlation $\rho_{\lambda\mu}$. We compare two estimators: two-way fixed effects (Time FE) and interactive fixed effects (IFE). The exhibits vary the panel length, the loading correlation and general-equilibrium feedback – the ingredients that the analysis of Section 3 identifies as the drivers of each failure mode; the appendix battery covers the remaining dimensions.

Experiment G2: loading correlation. Table 1 reports results for $\rho_{\lambda\mu} \in \{-0.5, -0.3, 0, 0.3, 0.5\}$, holding loading heterogeneity fixed at $\sigma = 1$. The Time FE estimates exhibit bias of the same sign as $\rho_{\lambda\mu}$ and increasing in its absolute value, consistent with Lemma 1. At $\rho_{\lambda\mu} = 0$, the estimator is approximately unbiased, but coverage remains below the nominal level, as also Experiment G5 will highlight more in detail. By contrast, IFE remains approximately unbiased and attains coverage close to 0.95 across all values of $\rho_{\lambda\mu}$.

Experiment G3: what the pooled estimators target when the responses differ. The experiments so far hold the response to the regressor constant across its components, so that a single slope describes the outcome completely. Once the outcome responds differently to synchronized and to idiosyncratic movement there is no such slope: α_c and α_x are both parameters of the data generating process, and an estimator reporting one number must combine them. Lemma 2 says how – each converges to $w_e \alpha_c + (1 - w_e) \alpha_x$, the weight w_e being the share of synchronized variation remaining after fixed effects are removed. Table 2 exhibits this directly, in the two-aggregate design of Experiment G6 and with uncorrelated

TABLE 1: Loading Correlation

Estimator	<i>Loading correlation $\rho_{\lambda\mu}$</i>				
	−0.5	−0.3	0	+0.3	+0.5
Time FE	0.753	0.853	1.000	1.145	1.245
	{0.00}	{0.00}	{0.29}	{0.00}	{0.00}
IFE (rank 1)	1.000	1.000	1.000	1.000	1.000
	{0.95}	{0.95}	{0.95}	{0.95}	{0.96}
<i>Condition: $\sigma = 1$. {Coverage}</i>					

Notes: 500 replications. $\sigma = 1$, $N = 250$, $T = 100$. IFE is approximately unbiased throughout. Generated by a separate simulation from Table 5, so the $\rho_{\lambda\mu} = 0$ column need not reproduce that table’s $T = 100$ cell exactly.

loadings, so that the bias channel of Lemma 1 is shut: the micro-local slope is held at $\alpha_x = 1$ while the response to the synchronized component α_c is raised from 1.0 to 2.5.

The estimates climb with α_c even though α_x has not moved – time fixed effects from 1.00 to 1.74, interactive fixed effects from 1.00 to 1.47 before falling back to 1.34. The estimates reflect the weighted averages implied by each specification: they are returning the blends the lemma predicts, at retained weights of 0.49 for time dummies and 0.65 for factor absorption at the true factor, and the time-dummy column reproduces $0.49\alpha_c + 0.51\alpha_x$. The discrepancy concerns the *interpretation* of the coefficient. For example, when $\alpha_c = 2$ and $\alpha_x = 1$, the mean Time FE estimate is 1.48, close to the weighted average implied by the lemma. Interpreting this coefficient as the response to idiosyncratic tightness would therefore overstate α_x by almost 50%. The pooled IFE estimator also departs from α_x as the two slopes diverge: coverage of α_x falls from 0.95 when $\alpha_c = \alpha_x = 1$ to zero when $\alpha_c = 2.5$ ²².

Experiment G4: the two-step procedure recovers the micro-local slope. Corollary 1(ii) establishes that defactoring the regressor recovers the micro-local slope without an instrument. Table 3 runs it on the same design: the regressor’s common component is extracted by principal components from the regressor panel alone, the loadings are fitted unit by unit, and the outcome is regressed on the idiosyncratic remainder. The estimator recovers α_x and is flat in α_c – 1.000 when the two responses coincide, 1.001 when the synchronized response is two and a half times the micro-local one, against the 1.34 the pooled estimator

²²For feasible rank-one IFE, the estimated factor increasingly captures the regressor’s common component as α_c rises. The weight on synchronized variation consequently declines, and the estimates fall below the weighted average calculated at the estimated factor. This pattern persists across the panel sizes considered, suggesting misspecification of the pooled rank-one model rather than a finite-sample discrepancy.

TABLE 2: Pooled estimators under $\alpha_c \neq \alpha_x$

α_c	Estimates of $\alpha_x = 1$		Predicted blend	Retained weight w_e		Coverage
	Time FE	IFE ($r = 1$)	IFE ($r = 1$)	Time FE	IFE	of α_x , IFE
1.00	1.000	1.000	1.000	0.49	0.65	0.95
1.50	1.244	1.274	1.314	0.49	0.63	0.02
2.00	1.485	1.472	1.575	0.48	0.57	0.01
2.50	1.736	1.338	1.541	0.49	0.36	0.00

Notes: Data generating process: $y_{it} = \alpha_c \mu_i F_t^x + \alpha_x \xi_{it} + \lambda_i F_t + \varepsilon_{it}$, $x_{it} = \mu_i F_t^x + \xi_{it}$, with $N = 250$, $T = 100$, $\alpha_x = 1$ and loadings drawn independently, so the bias channel of Lemma 1 is shut. Each estimator converges to $w_e \alpha_c + (1 - w_e) \alpha_x$ (Lemma 2). The IFE weight is computed at the estimated factor; at the true factor it is 0.65 in every row. The last column is the coverage of nominal 95% IFE intervals for α_x . $R = 500$.

delivers in the same cell – thereby eliminating the estimand conflation documented in Experiment G3. The three specifications differ in their treatment of the outcome’s common factor: they include the true factor, estimate it or omit it. All three recover α_x because the defactored regressor is orthogonal to both common factors. Omitting the outcome factor therefore does not bias the micro-local slope²³.

TABLE 3: The two-step procedure and the idiosyncratic slope

Estimator of $\alpha_x = 1$	Response to the synchronized component	
	$\alpha_c = 1.00$	$\alpha_c = 2.50$
Defactored (true factor)	1.000	0.999
Two-step, outcome factor not modelled	1.000	1.000
Two-step (baseline)	1.000	1.001

Notes: Same data generating process as Table 2, at zero measurement error. The micro-local slope is held at $\alpha_x = 1$ while the response to the synchronized component α_c takes the two ends of the range of Table 2. The two-step extracts one principal component from the unit-demeaned regressor panel and regresses the outcome on the idiosyncratic residual; the baseline additionally models the outcome’s factor structure, and the defactored row replaces the estimated factor with the true one. $R = 500$.

Experiment G5: identifying the synchronized response. Remark 2 establishes that a causal interpretation of α_c requires an instrument for aggregate variation. Table 4 evaluates this second step in a design where the common component of slack contains both instrumentable and confounded variation, $F_t^x = z_t + h_t$, with h_t entering the outcome through the

²³The rank selection this route needs is feasible: the eigenvalue-ratio and growth-ratio criteria applied to the regressor panel select the true single factor in every replication, so the diagnosis the procedure depends on can be made on the data a researcher holds.

loading $\theta\mu_i$. The resulting projection coefficient differs from α_c by $\theta \text{Var}(h)/\text{Var}(F^x)$ but, instrumenting the synchronized component with $\mu_i z_t$ removes this confounding. The feasible two-step estimator, which replaces the common component and its loadings with their principal component estimates, introduces only modest attenuation, of about two percent of α_c , reflecting the high accuracy of the first-stage factor extraction. Identification of α_c , however, remains aggregate: the instrument varies only over time, so the effective number of identifying observations is T , irrespective of the size of the cross-section. Appendix C extends the experiment to inference, examining instrument persistence and alternative variance estimators.

TABLE 4: Identifying the Synchronized Response

Design		Estimator of α_c		
α_c	θ	Projection	IV	IV, feasible
1.00	0.00	1.000	1.000	1.000
1.00	0.50	1.248	0.997	0.986
1.00	1.00	1.497	1.000	0.988
2.50	0.00	2.500	2.501	2.470
2.50	0.50	2.752	2.499	2.457
2.50	1.00	3.002	2.503	2.448

Notes: The aggregate slack impulse splits as $F_t^x = z_t + h_t$, independent AR(1) processes with $\text{Var}(h)/\text{Var}(F^x) = 0.5$; z is excluded from the outcome and h enters it with loading $\theta\mu_i$, so $\theta = 0$ is the no-confound benchmark and the projection’s plim is $\alpha_c + 0.5\theta$. Projection: mean OLS coefficient on the synchronized component c_{it} . IV: median coefficient instrumenting c_{it} with $\mu_i z_t$; feasible replaces μ_i and c_{it} by their principal component estimates. $R = 500$. Table A.3 in Appendix C reports means, coverage and the first stage.

Experiment G1: varying the panel length. Table 5 reports performance as T ranges over 20, 50, 100, 200, holding heterogeneity fixed ($\sigma = 1$) with loading correlation set to zero ($\rho_{\lambda\mu} = 0$). Consistent with Lemma 1, the mean Time FE estimate remains close to the true value of unity. Coverage, however, declines monotonically with T . Lemma 3 explains this pattern: the contribution of loading heterogeneity to the sampling variance remains of order $1/N$, whereas the heteroskedasticity-robust variance estimator is of order $1/(NT)$. The resulting confidence intervals therefore become increasingly narrow relative to the sampling dispersion of the estimator. Table 1 allowed for correlated loadings, introducing bias in the point estimate as well. IFE accounts for heterogeneous factor loadings and maintains coverage close to the nominal level in both experiments.

TABLE 5: Performance vs. Time Dimension

Estimator	<i>Time periods T</i>			
	20	50	100	200
Time FE	1.000 {0.64}	1.002 {0.49}	1.001 {0.37}	1.000 {0.24}
IFE (rank 1)	1.000 {0.95}	1.000 {0.94}	1.000 {0.94}	1.000 {0.97}
<i>Condition: $\rho_{\lambda\mu} = 0$ and $\sigma = 1$. {coverage}</i>				

Notes: 500 replications. DGP: $\alpha = 1.0$, $\rho_F = 0.80$, $\rho_\xi = 0.80$, $N = 250$, $\sigma = 1$.

Experiment G6: measurement error and the quality of the regressor. So far, the experiments document the limitations of Time FE. The final experiment turns the attention on IFE and asks what the estimator demands of the data. Because IFE identifies the slope from the variation in the regressor that survives the factor projection, classical measurement error – which, being idiosyncratic, survives every projection – occupies a larger share of the identifying variation than it does in a design that leans on the common component.

Table 6 quantifies the boundary in the configuration that matches the application: the regressor loads on its own aggregate G_t , distinct from the outcome’s factor F_t , the loadings are uncorrelated so that any movement is attributable to measurement error alone, and the noise share ranges from zero to 0.90. Both estimators attenuate, as any estimator that does not correct for errors-in-variables must. What is specific to IFE is that it attenuates *beyond* the classical benchmark: the estimated factor rotates away from the outcome’s aggregate toward the regressor’s ($|\text{corr}(\hat{g}_t, F_t)|$ falls from 0.999 to 0.789 while $|\text{corr}(\hat{g}_t, G_t)|$ climbs from 0.231 to 0.731). The mechanism is self-reinforcing: as noise attenuates the estimate, the unabsorbed signal $(\alpha - \hat{\alpha})\mu_i G_t$ remains in the residual, the factor rotates to absorb it, the projection strips more signal from the regressor, and the estimate falls further.

The result is a crossover, not a uniform ranking: up to a noise share of a quarter IFE remains the better-centered estimator, at one half the ranking flips, and beyond it the estimator this paper recommends is the worse one²⁴. Experiment G13 (Appendix C) repeats the exercise in the shared-factor design of Experiments G1–G2 and reinforces the conclusions:

²⁴Coverage of the true α is zero for both estimators at every positive noise share, and is reported for completeness rather than as a separate finding. Neither estimator is consistent for α under classical measurement error, and at $N = 250$ and $T = 100$ the sampling standard deviation is 0.012–0.053, so an attenuation of 0.10 places the truth more than ten standard deviations away: coverage of α records the attenuation with a ruler whose fineness is set by the sample size rather than by the estimator. The zero-noise column, where the comparison is informative, separates the two as Lemma 3 predicts – 0.94 for IFE against 0.76 for Time FE.

TABLE 6: Measurement Error in the Regressor

	<i>Noise share of the observed regressor</i>					
	0	0.10	0.25	0.50	0.75	0.90
Attenuation benchmark	1.000	0.900	0.750	0.500	0.250	0.100
Time FE	1.001 {0.76}	0.852 {0.00}	0.656 {0.00}	0.391 {0.00}	0.176 {0.00}	0.066 {0.00}
IFE (rank 1)	1.000 {0.94}	0.875 {0.00}	0.683 {0.00}	0.383 {0.00}	0.151 {0.00}	0.052 {0.00}
$ \text{corr}(\hat{g}_t, F_t) $	0.999	0.996	0.981	0.921	0.828	0.789
$ \text{corr}(\hat{g}_t, G_t) $	0.231	0.276	0.380	0.544	0.679	0.731
Share of x surviving projection	0.933	0.924	0.897	0.879	0.911	0.954

Notes: Data are generated from $y_{it} = \alpha x_{it} + \lambda_i F_t + \varepsilon_{it}$ and $x_{it} = \mu_i G_t + \xi_{it}$, with the regressor observed with classical error, $x_{it}^{\text{obs}} = x_{it} + \nu_{it}$, and the noise share defined as $\text{Var}(\nu)/\text{Var}(x^{\text{obs}})$. Coverage of the true $\alpha = 1$ by nominal 95% confidence intervals, in braces. The attenuation benchmark is $\alpha \cdot \text{Var}(x)/\text{Var}(x^{\text{obs}})$, the attenuation a textbook errors-in-variables correction would predict. $\hat{g}_t$ denotes the estimated rank-one factor; the last row reports the share of the within variance of the observed regressor that survives the estimated-factor projection. $R = 500$.

with no separate aggregate to retain, IFE identifies from roughly a third of the variation Time FE keeps – precisely where classical noise concentrates – and attenuates below both the benchmark and Time FE at every noise level.

The boundary is therefore quantitative and suggests some robustness to measurement error: the estimator’s advantage over time dummies survives until noise accounts for roughly *half* of the regressor’s unit-level variance. What is needed is a measure of slack that carries signal in the cross-section at the frequency of estimation, not merely one correlated with tightness on average – the standard against which the next section builds ours.

A further set of experiments – reported in Section 7.6 and Appendix C – confirms that the results are robust to alternative variance estimators, calibrations and factor ranks.

6 Measuring Slack: The PLS Indicator

The estimator pins down the slope using the idiosyncratic variation in slack that is left after removing the common factors, so the slack measure must preserve informative variation along that dimension – granular, observed at the estimation frequency and separable from the common component (Assumption 4, Experiment G6). This section introduces a measure that meets this requirement, sets out its construction and reports validation evidence.

Despite the little agreement on the appropriate empirical measure of slack²⁵, a growing number of recent papers have proposed the vacancy-to-unemployment (V/U) ratio²⁶. While the V/U ratio could be a promising candidate, it is impractical in the euro area context mainly due to the unavailability of accurate (and granular) data²⁷. For this reason, we propose a granular indicator of labor market tightness that reflects firms' perceived hiring difficulties and serves as an alternative to the vacancy-to-unemployment ratio.

Our primary data source is the Joint Harmonised EU Programme of Business and Consumer Surveys (BCS), which is administered by the European Commission's Directorate-General for Economic and Financial Affairs each month, with additional quarterly questions. The surveys encompass a large sample of firms: 19,380 firms for the 24 manufacturing sub-sectors for the countries in our sample. We use data for ten euro area countries (Austria, Belgium, Germany, Greece, Spain, Finland, France, Italy, the Netherlands and Portugal²⁸) and the baseline sample period spans from 2002 to 2024²⁹.

Participation is mandatory in several of the countries we consider (Spain, France, Italy, the Netherlands and Portugal), ensuring high response rates and data quality. According to the European Commission's methodological guide, the sample of firms does not differ much from one period to another, as a panel of companies is established and surveyed each month and quarter.

Among the questions asked in the surveys, we focus on responses to the quarterly *Question 8*, which asks firms “*What main factors are currently limiting your production?*”. Respondents are requested to select one or several factors: (i) none; (ii) insufficient demand; (iii) shortage of labor force; (iv) shortage of material and/or equipment; (v) financial constraints; (vi) other factors. The BCS dataset then reports, at the NACE-2 level (i.e. at the sub-sector level for each country)³⁰, the share of firms that respond that their production is currently

²⁵See [Albuquerque and Baumann \(2017\)](#) and [Hong et al. \(2018\)](#).

²⁶See, e.g., [Barnichon and Shapiro \(2024\)](#); [Benigno and Eggertsson \(2023\)](#); [Ball et al. \(2022\)](#).

²⁷For instance, vacancy data are generally not available even at the aggregate (country) level, with some of the main euro area countries missing entire time series.

²⁸We exclude Luxembourg, for which inflation (PPI) is missing, and Ireland, for which PLS data are only available for the most recent period.

²⁹While data are available for some countries and sub-sectors from 1990Q1, in the baseline results presented we restrict the sample starting from the first quarter of 2002 to avoid capturing co-movements due to the convergence period before the introduction of the euro. Results are robust to this choice and the table for the full 1990-2024 sample is reported in the Appendix.

³⁰NACE is the French acronym for “*Nomenclature statistique des Activités économiques dans la Communauté Européenne*” and is the statistical classification of economic activities used by the European Union to ensure comparability of data across member states. See Table [A.14](#) in the Appendix for more details on the sub-sectors considered in this study.

constrained by each of those factors³¹.

TABLE 7: Correlation of PLS with Standard Labor Market Measures

Euro Area						
Aggregation Level	V/U	Unemp.	Vac.	$1 - u$	ULC	Lab. Share
Country	0.896	-0.376	0.587	0.506	0.276	0.212
Euro Area	0.963	-0.900	0.876	0.846	0.378	0.044
United States						
U.S.	0.912	-0.531	0.960	0.784	-0.007	-0.466

Notes: Pearson correlation coefficients between PLS and each measure. Euro area aggregates use official EU Commission country weights; country rows are sub-sector-weighted country panels. Unemp.: number of unemployed persons. Vac.: job vacancies. $1 - u$: one minus the unemployment rate. ULC: nominal unit labour costs, % change year-on-year; for the euro area, per hour worked; for the U.S., nonfarm business sector. Lab. Share: for the euro area, compensation of employees as a share of GDP at current prices; for the U.S., the nonfarm business labor share index. Euro-area sample: 2002Q1–2024Q4. U.S. sample: 1974Q1–2024Q4.

Table 7 (top panel) reports the correlation between PLS and the vacancy-to-unemployment (V/U) ratio as well as with its separate components (vacancies and unemployment) for those countries for which Eurostat data are available (Germany, Netherland and Portugal). Although these series are constructed at the country level, we also compute correlations at the aggregate euro area level³². The key result is that PLS is strongly and positively correlated with the V/U ratio, with a correlation coefficient close to one. As anticipated, PLS is positively correlated with vacancies and negatively correlated with unemployment, which supports its use as a proxy for labor market tightness. Figure 1a illustrates the joint movement of the V/U ratio and PLS. We also benchmark PLS against the employment rate ($1 - u$), nominal unit labor costs (ULC) and the labor share.

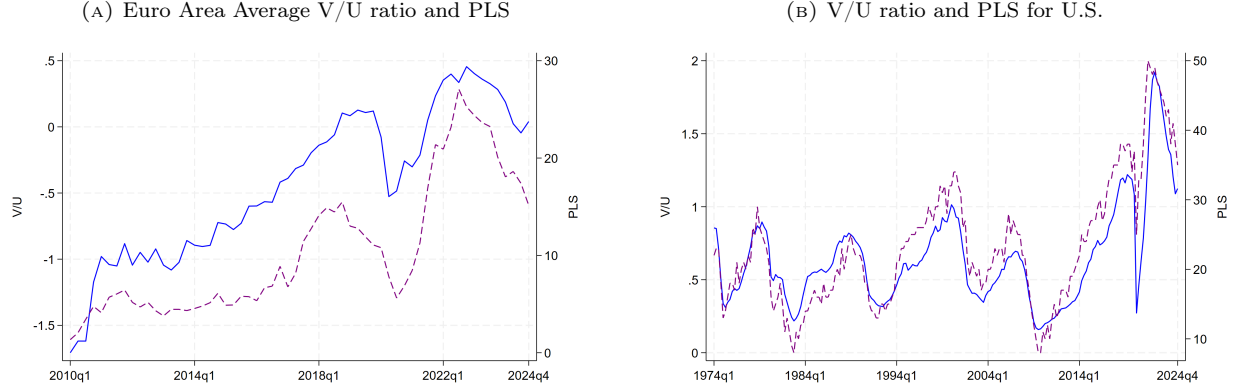
We corroborate this relationship further by showing that a U.S. survey measure closely related to PLS is likewise strongly positively correlated with the V/U ratio and with $1 - u$. These results are presented in Table 7 (bottom panel) and in Figure 1b. Details on data sources are provided in the Appendix³³.

³¹Some studies use the survey measures to approximate vacancies or analyze aggregate wage-setting (Kiss et al., 2022; Bonthuis et al., 2016; Bonam et al., 2021); our approach exploits sub-sector-country granularity to link labor shortages to inflation dynamics.

³²Throughout, “euro area” denotes the set of countries in our sample, as defined earlier and in the Appendix.

³³We thank Pierpaolo Benigno for kindly providing his U.S. vacancies and unemployment data.

FIGURE 1: Correlations of V/U and PLS



Notes: The figures show the time-series of aggregate PLS – in purple (dashed), right scale – and the natural logarithm of the V/U ratio – in blue, left scale. The log is used for readability; the correlations in Table 7 are computed on the level of V/U (for the euro area aggregate, the level and log correlations are 0.96 and 0.92 respectively). The plots start at different points in time due to data availability of the V/U series. When aggregating the euro area data, we compute weighted averages based on official weights reported by the EU commission. U.S. data are downloaded directly at the country level.

The strong correlation with the vacancy-to-unemployment ratio supports interpreting PLS as a measure of labor market tightness. In New Keynesian models, pricing decisions depend on real marginal costs (e.g., Galí, 2015). Search-and-matching frictions provide a channel through which hiring difficulties affect these costs: tighter labor markets reduce the probability of filling a vacancy and increase the expected recruitment costs required for an additional employee (e.g., Pissarides, 2000). Hiring costs can therefore raise the marginal cost of production even when wages remain unchanged, generating inflationary pressure.

Importantly, the correlations with the cost-side measures are consistent with this interpretation. The labor share captures *average* rather than marginal labor costs and is influenced by forces unrelated to hiring frictions, including bargaining power and institutional characteristics; PLS is close to uncorrelated with it. The moderate positive correlation with euro area nominal ULC growth is what one expects if tight labor markets eventually pass into measured labor costs. counterpart is zero. PLS is therefore not a stand-in for these indicators: it captures something ahead of and distinct from average measured costs, the vacancy-filling friction itself. This distinction justifies using PLS, rather than traditional cost-based indicators, as the slack measure in the Phillips curve estimation that follows.

Unlike some studies for the U.S., our dataset is based on harmonized questionnaires and a common methodology, which ensures comparability across countries and sub-sectors, pro-

viding a more robust framework for cross-country analysis.

Taken together, the evidence speaks to the requirement with which this section opened. PLS tracks the vacancy-to-unemployment ratio closely where that ratio can be constructed, so it carries the tightness signal; it is available at the sub-sector-country level at quarterly frequency, so the signal is present in the cross-section; and it is close to uncorrelated with the labor share and only moderately correlated with unit labour cost growth, so it is not a relabeling of the average-cost measures – a property that matters for identification, since a measure moving with the common component of inflation would retain little identifying variation once the factor is projected out.

7 Re-estimating the Phillips Curve

This section brings the preceding contributions together. We re-estimate the euro area Phillips curve, first assessing how far the time fixed effects designs that dominate the recent cross-sectional literature distort the estimated slope, then applying the interactive fixed effects estimator of Section 4, and doing so with the measure of slack developed in Section 6, whose properties the theory of Section 3 shows to matter. Throughout, the analysis executes the implementation checklist of Remark 4 step by step and we flag each step where it is performed.

The euro area sharpens this tension and is therefore a natural testing ground. Here the cross-sectional strategy has been hard to carry through, and estimates remain aggregate or country-level (Eser et al. 2020; Florio et al. 2025); at the same time the common shocks are unusually strong and unusually unequal in their reach – a single monetary policy, common energy shocks and exchange-rate movements, but sectors and member states that differ sharply in how each transmits to local prices and slack. This is precisely the environment in which time fixed effects fail and interactive fixed effects are identified. We therefore bring two ingredients to the euro area: a *granular* measure of labor market slack that opens up the cross-section and interactive fixed effects that absorb common factors through heterogeneous loadings rather than a common intercept. The results provide empirical support for the paper’s main point: the euro area Phillips curve is significantly steeper than conventional time fixed effects imply and the correction is economically large.

7.1 Data and Measurement

The slack measure and its construction are the subject of Section 6; here we describe the outcome and control data. Our inflation measure comes from Eurostat’s Short-Term Statistics database for the Producer Price Index (PPI). Inflation is measured as the year-on-year percentage change in the sub-sector’s domestic producer price index, at quarterly frequency, in percentage points. The PPI is particularly well-suited for our analysis as it aligns directly with BCS variables – particularly our measure of labor market slack – at the NACE-2 (sub-sector-country) level, allowing for a precise examination of the relationship between labor shortages and inflation.

Moreover, recent contributions (Reis, 2026; Beaudry et al., 2025) emphasize that Phillips curve estimates can be biased if short-term inflation expectations are not adequately accounted for. Therefore, in our analysis we will augment our most preferred specification with information on short-term selling price expectations from the BCS database. In particular, the BCS provides answers to the following question at the sub-sector-country level: “How do you expect your selling prices to change over the next 3 months?”. Firms can answer to the question in three ways: (i) increase, (ii) stay the same or (iii) decrease. We will include the responses in the regression analysis and confirm the robustness of our results.

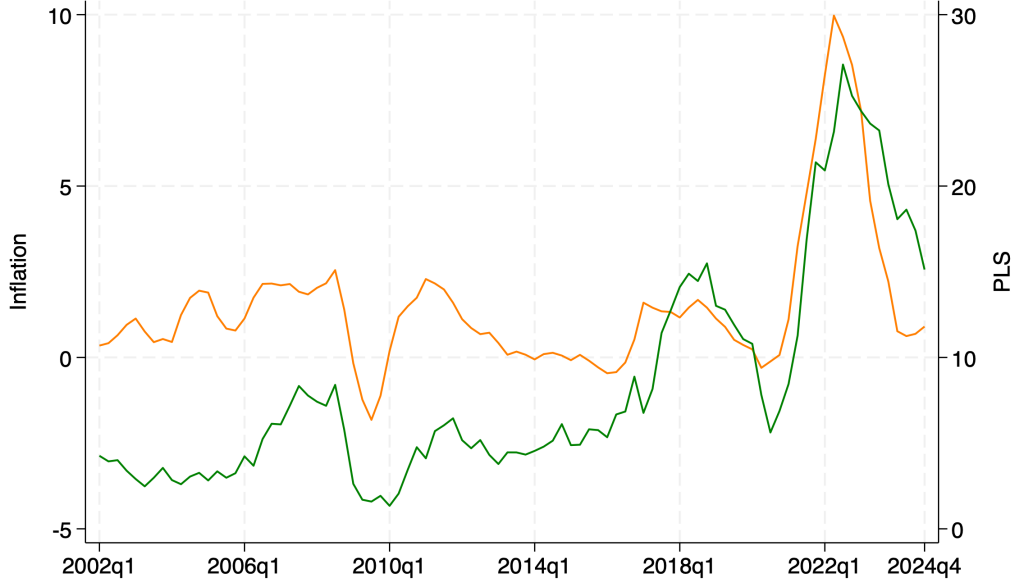
Appendix D.1 provides more detailed information on the data sources and the sample selection procedures adopted.

7.2 Estimation

Descriptive evidence. The weighted averages of both inflation and PLS are plotted in Figure 2 and show how the time series evolve over the sample period. It highlights two aspects of the data: (i) there is a clear co-movement between PLS and inflation and (ii) post-COVID, inflation and perceived labor shortages both rose sharply, pointing to heightened inflationary pressures and tighter labor markets in the wake of the pandemic.

Figure 3 shows the evolution of PLS for three sub-sector-country pairs – machinery in Germany, wood products in France and wearing apparel in Italy – over our sample period. There is certainly a great deal of comovement, with episodes of acute labor market tightness. Shortages rose sharply in 2007–08, again in 2018–19 and after the pandemic. Crucially though, the figure illustrates well that there is also substantial cross-pair variation. One example is the 2018–19 period, during which German machinery seemed to be tighter than French wood products, while Italian apparel producers reported only few labor shortages. Even

FIGURE 2: Euro Area Average Perception of Labor Shortages (PLS) and Inflation



Notes: The figure shows the time-series of aggregate PLS – in green, right scale – and (PPI) inflation – in orange, left scale. When aggregating, we compute weighted averages based on official weights reported by the EU commission.

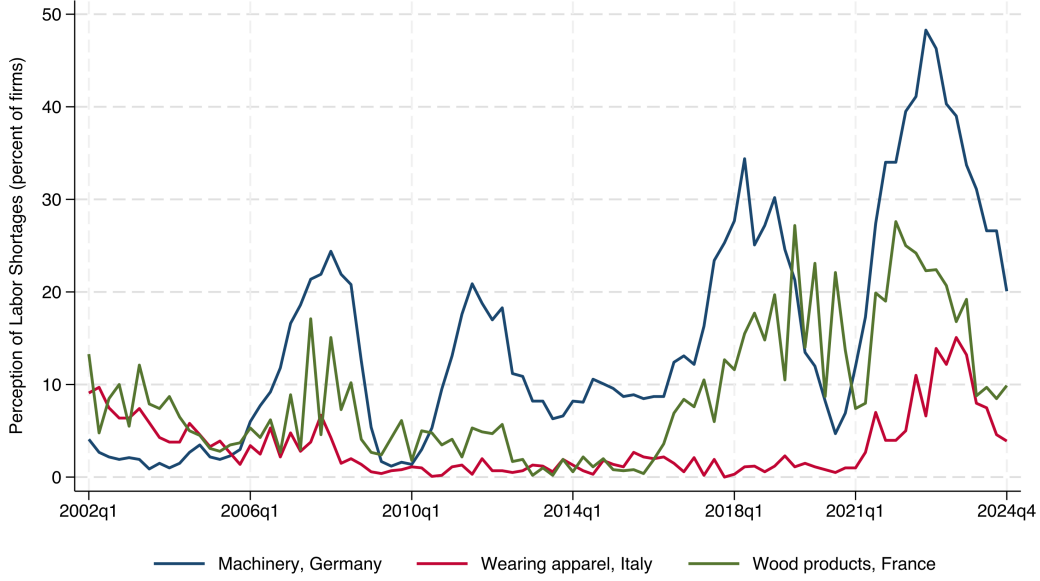
the post-pandemic surge, which reached all three, did so at very different amplitudes and horizons. Online Appendix [OA.2.1](#) is even more comprehensive in showing the considerable heterogeneity present in both the *timing* and the *intensity* of these episodes. This cross-sectional variation is particularly valuable in the context of the euro area. Unlike in the United States, the euro area is a recent monetary union. As a result, identifying the Phillips curve from aggregate time series alone becomes more challenging. Exploiting cross-country and sectoral heterogeneity provides a richer basis for estimating the slope of the Phillips curve in the euro area and allows us to make better use of the few observable episodes of labor market tightness.

Specifications. We estimate a sequence of nested models. The baseline (Table [8](#), column (1)) starts from

$$\pi_{ijt} = \alpha x_{ijt} + \gamma_i + \tau_j + \epsilon_{ijt}, \quad (13)$$

then adds time fixed effects ω_t , observable controls – Γ_{ijt} collecting the variables measured at the sub-sector-country level and Θ_{jt} those measured at the country and euro area level – and, finally, short-term inflation expectations. Time fixed effects (ω_t) absorb common shocks

FIGURE 3: Perceived Labor Shortages in Three Sub-Sector-Country Pairs



Notes: PLS for three sub-sector-country pairs, 2002Q1–2024Q4, quarterly.

by imposing homogeneous loadings – the source of bias characterized in Lemma 1. However, sectors and countries differ in many respects, including labor productivity, wage-setting institutions, bargaining power and labor market regulations. Hence, in column (3) we relax that assumption and use observable controls instead of time fixed effects³⁴. Controls are designed to capture time-varying forces that are plausibly correlated with inflation and labor market tightness. Replacing a single, unrestricted time dummy with economically interpretable aggregates – several of them measured at the sub-sector-country or country level – makes the specification *more flexible*, as it targets specific common forces rather than absorbing all common variation irrespective of its source. The validity of this approach depends on the extent to which observable controls capture latent determinants of inflation. Incomplete adjustment leaves a common component in the error term, biasing the estimated Phillips curve slope if that component remains correlated with labor market tightness conditional on the included controls.

³⁴Sub-sector-country specific controls available from BCS are: insufficient demand; shortage of material and/or equipment; financial constraints; other factors and a proxy for short-term inflation expectations. We also include deviations of the sub-sector-country relative producer price index from its steady-state value of one and, in the observable-controls specifications of columns (3) and (4) only, average inflation over the previous four quarters. No lagged inflation enters any interactive fixed effects specification, so the exogeneity condition mapped to Bai’s framework in Appendix B.6 – which rules out lagged dependent variables – holds for every reported IFE estimate. At the country level we include real effective exchange rates and energy prices, at the euro area level.

Treating inflation expectations. The treatment of inflation expectations is a central specification choice in cross-sectional Phillips curve estimation. Recent work shows that the resulting slope estimates can be particularly sensitive to how expectations are controlled for (Beaudry et al., 2025). In a panel at this level of disaggregation, absorbing expectations through period dummies imposes the restrictive assumption that their contribution to inflation is identical across units at each date. This restriction may be particularly strong across sectors and countries with different production structures and exposure to common shocks. For example, selling-price expectations may differ substantially between German chemical producers and Portuguese apparel manufacturers. The BCS provides a proxy for short-term inflation expectations at the sub-sector-country level, allowing us to control for expectations at the same level of disaggregation as inflation and labor shortages, consistent with the heterogeneity described in Section 3.

We therefore take the specifications including this proxy – column (4) of Table 8 and column (8) of Table 9 – as our preferred specifications. The results indicate that the sensitivity to these controls differs across estimators. Including the controls and expectations substantially reduces the Time FE slope, consistent with Reis (2026) and Beaudry et al. (2025). By comparison, the component slopes estimated under IFE decline only modestly and remain statistically significant.

The preferred specification is the IFE regression

$$\pi_{ijt} = \alpha x_{ijt} + \lambda_{ij} F_t^y + \gamma_i + \tau_j + \delta \Gamma_{ijt} + \eta_{ijt}. \quad (14)$$

Here λ_{ij} and F_t^y – the common component of inflation, in the two-aggregate notation of Section 3.2 – are estimated jointly with α by minimizing the least-squares objective over the slope, the factors and the loadings. The factor rank r , i.e. the number of common factors, is chosen using data-driven information criteria (Bai and Ng, 2002; Ahn and Horenstein, 2013; Onatski, 2010; Gagliardini et al., 2019)³⁵. Reassuringly, the estimated factor displays a profile consistent with common inflation dynamics and with economic intuition, troughing during the Great Financial Crisis and spiking post-pandemic (Figure A.1, Appendix D.5).

Given the importance of the granularity assumption, we verify it directly – Step 3 of the checklist in Remark 4. The equal-weighted HHI is 0.0081; under official EU Commission output weights, it is 0.022, with the largest pair carrying 6.2% of economic weight. The sharper condition the theory needs, $N \cdot \text{HHI} = O(1)$ (Appendix B.6), is comfortably met:

³⁵A more detailed discussion on the estimation procedure is provided in Online Appendix OA.1.3.

the product is 1.00 under equal weights and roughly 2.7 under output weights. The leave-top- k diagnostic passes: dropping the 1, 5, 10 or 24 largest pairs by economic weight and re-running the full two-stage construction leaves the estimated common factor of slack essentially unchanged – its correlation with the full-sample factor stays above 0.99 throughout – so no pair is large enough for its own variation to construct the aggregate (Appendix D.7, Tables A.20 and A.21). What these diagnostics establish is that the quantity the theory says matters – $|\psi| \cdot \max_i s_i$, mapped quantitatively by Experiment G12 – is small enough that residual general-equilibrium feedback cannot account for the results.

7.3 Empirical Findings

Tables 8 and 9 report estimates of the Phillips curve slope across different specifications.

TABLE 8: Phillips Curve Estimates

	(1) FE	(2) Time FE	(3) Controls	(4) Time FE + Infl. Exp.
PLS	0.157*** (0.015)	0.050*** (0.014)	0.043*** (0.006)	0.016** (0.008)
Observations	9094	9094	9094	9094

Notes: Column (1) reports the two-way (sub-sector and country) fixed effects specification of Eq. (13). Subsequent columns cumulatively add time fixed effects (2), control variables (3), and a proxy for short-term inflation expectations (4). Standard errors clustered by sub-sector–country pair are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fixed effects. We begin with the benchmark specification used in the literature. Column (1) of Table 8 includes sub-sector and country fixed effects, absorbing persistent structural differences across units; the estimated slope is sizable and statistically significant. Column (2) adds time fixed effects – the standard treatment of unobserved common forces, such as energy prices and inflation expectations – and the slope falls sharply.

Part of this fall is the restriction the dummies impose: they absorb common shocks through a single period mean, forcing every sub-sector and country to respond identically – exactly the homogeneity that Section 3 shows can be violated in disaggregated data. Column (2) thus removes confounding and genuine signal together, and the coefficient it reports partly reflects the estimator rather than the curve.

Controls. Column (3) replaces the time fixed effects with observable control variables, while retaining sub-sector and country fixed effects. The estimated Phillips curve slope is

smaller than in the two-way fixed effects specification in column (1), suggesting that part of the variation previously attributed to labor market tightness may reflect omitted factors correlated with PLS. At the same time, the coefficient remains larger than in the time fixed effects specification in column (2), indicating that time fixed effects may be overly restrictive and absorb variation related to labor market conditions. However, the result may reflect the limitations of control-based approaches, since observable proxies are imperfectly measured and may absorb variation that is instead attributable to labor market slack.

Column (4) provides the most stringent specification. It restores the time fixed effects of column (2) alongside the controls in column (3) and adds a proxy for short-term inflation expectations. The estimated slope falls to 0.016, roughly one third of its value before conditioning on expectations, consistent with Reis (2026) and Beaudry et al. (2025). Standard errors are clustered by unit to account for the loading-dispersion component of uncertainty characterized in Section 3.3.

TABLE 9: Phillips Curve Estimates

	(5) IFE	(6) FE	(7) Sec.×Cty	(8) Infl. Exp.
Common (α_c)	0.163*** (0.010) [0.021]	0.153*** (0.010) [0.023]	0.149*** (0.010) [0.023]	0.124*** (0.009) [0.020]
Idiosyncratic (α_x)	0.039*** (0.006) [0.013]	0.058*** (0.009) [0.020]	0.052*** (0.009) [0.020]	0.043*** (0.007) [0.016]
Observations	9094	9094	9094	9094

Notes: PLS is decomposed in two stages before estimation. α_x is the micro-local slope; α_c is a projection. Columns add cumulatively: (5) interactive fixed effects; (6) sub-sector and country effects; (7) sub-sector $\times$ country loadings; (8), the featured specification, a proxy for short-term inflation expectations. Robust standard errors in parentheses; bootstrap standard errors in brackets, from 200 block-bootstrap replications over pairs that re-run stages 1 and 2 within each. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, from the bootstrap standard errors.

Interactive fixed effects. Columns (5)–(8) report the estimator Section 4 builds for exactly such an environment: the latent factor and its unit-specific loadings are estimated jointly with the slope, so the common component is removed from both inflation and slack before the coefficient is estimated – with valid inference, under the granularity condition verified in Section 7.2 (Proposition 1).

The estimation follows the two-step estimator of Remark 2 and each stage corresponds to

a step of the checklist. Following *Steps 1 and 4*, we diagnose the factor structure: the first principal component of pair-demeaned PLS accounts for roughly two thirds of its within variance (Figure A.3), the rank criteria select one factor (Table A.17, Appendix D.4), and the extracted series correlates only 0.305 with the estimated inflation factor – the two-aggregate case, in which a pooled slope blends two estimands (Lemma 2). Following *Step 5*, we therefore defactor: PLS is split into its synchronized component $\mu_{ij}F_t^x$ and its idiosyncratic remainder, and both slopes are estimated jointly – α_x is the micro-local object, α_c the common-component projection, whose causal interpretation requires an additional identifying restriction. Because the factor and loadings are generated regressors, the table reports, in square brackets, standard errors from a block bootstrap over pairs that re-runs both stages inside every replication, and the stars are computed from them (Appendix D.9 discusses more in detail).

Columns (5)–(7) implement the component decomposition. Across specifications, the synchronized slope remains substantially larger than the idiosyncratic one. Adding sub-sector and country effects increases the estimated idiosyncratic slope. Allowing factor loadings to vary at the sub-sector-country level modestly reduces both component slopes, while the synchronized slope remains substantially larger³⁶. The implied omitted variable bias from Lemma 1 is small, about -0.005 , so the main distinction is between the synchronized and idiosyncratic components rather than a large bias correction.

Column (8), our baseline specification, adds the sub-sector-country proxy for short-term inflation expectations. The distinction between the two components remains pronounced: the synchronized slope is 0.124, against 0.043 for idiosyncratic tightness, and both remain statistically significant. Conditioning on expectations lowers both coefficients modestly, in the direction anticipated by Reis (2026) and Beaudry et al. (2025). However, the contrast with the conventional time FE specification is particularly marked. In fact, adding the same expectations control reduces the time-FE slope to 0.016, whereas under the factor specification the synchronized and idiosyncratic slopes remain 0.124 and 0.043, respectively. Thus, conditioning on expectations substantially changes the benchmark estimated used in the literature, without eliminating either component of the Phillips curve. This comparison is possible because the expectations proxy varies across sub-sectors and countries; a common series would be absorbed by the estimated factor and could not provide the this test. We therefore take column (8) as the baseline specification.

³⁶The estimated loadings confirm the heterogeneity: their cross-sectional distribution is markedly right-skewed and departs visibly from normality (Figure A.2, Appendix D.5). The dispersion is mostly a sub-sector phenomenon: euro area countries share a currency and broadly similar institutions, while sub-sectors differ in technology, price flexibility, unions and market structure.

7.4 Why the Estimators Differ: Composition and Magnitudes

Lemma 2 (Failure II) establishes that, when the outcome responds differently to the synchronized and idiosyncratic components of the regressor, each estimator converges to a weighted average of α_c and α_x . The corresponding weight is determined by the share of synchronized variation that remains after the estimator’s absorption step. Since both the component slopes and the associated weights are estimable, the decomposition can be implemented empirically and the this subsection reports that decomposition.

The empirical results suggest that this distinction is quantitatively important and splitting tightness into its two components gives two very different slopes. Inflation in a sector-country pair rises by 0.124 for a unit of synchronized tightness – the euro area cycle, scaled by the pair’s exposure – and by 0.043 for a unit of tightness specific to the pair. The two pooled estimates are averages of these slopes and the averaging weights explain their difference. IFE draws 48 percent of its identifying variation from the synchronized component, while time fixed effects draw 28 percent. Less, but not zero: period dummies remove the euro area *average* of the cycle, not the cycle’s unequal footprint across pairs, so part of the synchronized component survives the demeaning.

Why do the two pooled estimators differ? Table 10 answers by decomposing each slope in three parts: the blend of the component slopes at the estimator’s own weight, the omitted variable term of Lemma 1 and the co-movement of inflation with the quarter averages of tightness that the estimator removes. The decomposition is evaluated using a parsimonious specification without observable controls, corresponding to the setting considered in the lemmas.

We highlight three findings. First, the estimand shift of Lemma 2 – Failure II – accounts for about forty percent of the gap between the two slopes: the estimators differ in part because they average the same two slopes with different weights. Second, the omitted variable term – Failure I – is an order of magnitude smaller; the textbook channel is not what separates the estimators here. Third, period dummies remove the quarter-to-quarter movement of euro area tightness, and inflation co-moves with it. That lost co-movement pulls the time fixed effects estimate below the average of the two component slopes it targets, whereas IFE retains the synchronized variation and therefore places full weight on α_c . Deleting the period averages therefore imposes two costs on the time fixed effects design: it alters the estimand and it removes from the identifying variation the co-movement of inflation with euro area tightness.

TABLE 10: The gap between the estimators

	Time fixed effects	IFE
Estimated slope on PLS	0.048	0.097
Blend of the component slopes, $w_e \alpha_c + (1 - w_e) \alpha_x$	0.084	0.104
Omitted-variable term (Lemma 1)	0.006	-0.006
Co-movement with the quarter averages the estimator deletes	-0.042	-0.001

Notes: Clean specification, no controls. Each column decomposes the estimated slope exactly into the blend of the component slopes ($\alpha_c = 0.151$, $\alpha_x = 0.059$) at the estimator’s weight on the synchronized component ($w_{\text{TFE}} = 0.278$, $w_{\text{IFE}} = 0.488$), the omitted variable term of Lemma 1 and the co-movement with the quarter averages of PLS the estimator deletes.

The deleted co-movement can be reconciled in two different ways. The first is general equilibrium. Consider a multi-region model with a single currency. Expectations, wage norms and the policy stance are set at the union level, so a region’s inflation responds to aggregate tightness, alongside its own. A regional regression with period dummies differences that response out – this is what makes it a regional curve (Hazell et al., 2022) – and it is the response the missing-intercept literature identifies as unrecoverable from cross-sectional designs (Wolf, 2023; Moll and Hanney, 2025). The second source is the persistence of the pressure itself. In a forward-looking Phillips curve the slope on current tightness rises with the expected persistence of the pressure, since firms reprice for cost pressure that will last rather than for pressure that is transitory. The euro area cycle in PLS is highly persistent; the idiosyncratic component reverts within a quarter. Under either mechanism, period dummies absorb the component of tightness to which firms’ pricing decisions are most sensitive to.

However, we have to interpret these estimates with some caution. The weight on the synchronized component is a property of the sample’s variance composition, not an economic parameter; and it is the idiosyncratic slope, not the synchronized one, that corresponds to the micro-local object of Section 3. The synchronized slope itself admits two readings. It may be a genuinely stronger response of inflation to synchronized tightness – the wedge the missing-intercept literature argues cross-sectional designs cannot recover (Wolf, 2023; Moll and Hanney, 2025) – or it may partly reflect aggregate forces that move with tightness and enter inflation directly, in which case α_c is an upper bound.

Separating the two interpretations, therefore, requires an instrument for aggregate slack. The external instruments do not satisfy relevance and exclusion restrictions: high-frequency monetary policy surprises (Altavilla et al., 2019) move euro area tightness only through their

information component, which enters inflation directly, while the policy component has a weak first stage (Appendix D.6). We therefore construct the instrument inside the panel, following Sarto (2025): the innovation to the aggregate movement in PLS that survives once each pair’s response to its own common component of inflation is removed. Its validity rests on one restriction, that this innovation is orthogonal to the common shocks entering inflation directly. Standard errors are clustered by time (Experiment G5) and the first-stage F stat is above ten. In the featured specification, the IV estimate of α_c is 0.19, which is significant and robust to weak-IV confidence bands (Table A.18). The estimate is larger than the projection coefficient, equal to 0.124, and is approximately four times the estimated micro-local slope α_x . Therefore, α_c should be interpreted as the projection in the tables and the IV should be read as its causal counterpart.

Magnitudes. Standardized, a one standard deviation movement in a pair’s own idiosyncratic tightness raises its inflation by a quarter of a point; a one standard deviation synchronized movement raises it by 1.3 points on the IV slope. During the recent inflation episode, this distinction is economically substantial. Between the trough in 2020Q3 and the peak in 2022Q3, the average share of firms reporting labor shortages rose by 18.1 percentage points, 17.4 of them through the synchronized component; the average idiosyncratic component was essentially unchanged. Applying the IV estimate of α_c to the synchronized component implies that it accounts for about 3.3 percentage points of the 12.6-point increase in producer price inflation. The remaining increase is attributed to the common forces captured by the outcome factor. The implied slopes also differ sharply in economic terms: reducing inflation by one percentage point requires a decline in synchronized shortages of about five percentage points, roughly one standard deviation of their aggregate variation, compared with about twenty-three percentage points through idiosyncratic tightness, or roughly four standard deviations. The Phillips curve is therefore substantially steeper with respect to synchronized than to idiosyncratic tightness.

The decomposition shows whether identification comes from synchronized or idiosyncratic variation. This distinction is central for the non-linear analysis in the next section.

7.5 Non-Linearity and the Location of the Identifying Variation

We next examine whether the Phillips curve steepens as labor markets tighten, a question central to recent debates on inflation dynamics. A prominent strand³⁷ argues that the curve is state-dependent, steepening at high levels of labor market tightness; Benigno and

³⁷See, among others, Faber and Züllig (2026); Gitti (2024); Babb and Detmeister (2017); Murphy (2017).

Eggertsson (2023), in particular, provide theoretical and empirical evidence for the U.S. of an inverse-L shape, with inflation responding little to slack in loose labor markets but sharply once tightness passes a certain threshold. We test this framework with disaggregated euro area data, our granular slack proxy and the extended interactive fixed effects estimand.

Which component carries the non-linearity? Does the curve steepen because a sub-sector’s *own* labor market is tight, or because *everyone’s* is? The two answers point to different economics. In the first, the mechanism is local: a firm that cannot hire faces rising marginal costs and raises prices, whether or not the rest of the economy is tight – so idiosyncratic tightness should steepen the curve just as synchronized tightness does. In the second, the mechanism runs through the aggregate state: firms reprice more readily when many reprice at once and expectations and wage norms respond to broad-based inflation rather than to one industry’s hiring troubles – so the steepening should appear only in the synchronized component, and only when the economy-wide labor market is tight. The question is empirically relevant because the tight episodes and the synchronized variation coincide: Figure 4 plots the estimated common factor of PLS against the share of sub-sector-country pairs above their tightness threshold, and the two series move together (correlation 0.92), peaking jointly in 2007, in 2018–19 and post-pandemic. The concentration is quantitatively relevant: observation-quarters above threshold account for 38 percent of the sample but 71 percent of the within variance of the common component of PLS – the variation whose average time fixed effects delete is disproportionately the variation of the tight episodes³⁸.

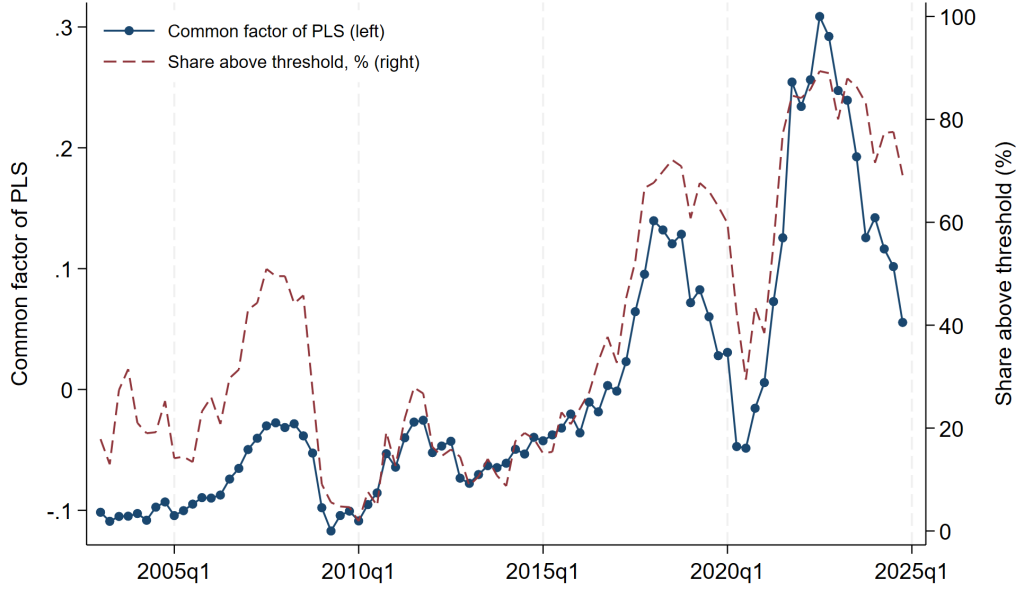
We test directly which of the two separate readings is supported in the data. Let S_t be the share of sub-sector-country pairs above their tightness threshold at date t – the series of Figure 4 – and let $D_t = \mathbb{1}\{S_t \geq s^*\}$ mark the tight aggregate state. We then interact both components of slack with D_t in the baseline specification,

$$\pi_{ijt} = \alpha_c \mu_{ij} F_t^x + \alpha_x \xi_{ijt} + (\gamma_c \mu_{ij} F_t^x + \gamma_x \xi_{ijt}) D_t + \phi D_t + \delta \Gamma_{ijt} + \lambda_{ij} F_t^y + \alpha_i + \tau_j + \eta_{ijt}. \quad (15)$$

Rather than imposing a common break point, we estimate the thresholds for each sub-sector-country pair with a model-free random forest procedure that lets the data locate where the inflation-slack relationship changes slope; the implementation, a simulated validation, the same test run with a pair-level rather than an aggregate state, an alternative-threshold

³⁸This is a fact about where the synchronized movement lives, not a second channel for the gap: the two estimators draw on the steep region almost equally, so the threshold partition accounts for under two percent of the gap between the pooled estimates – the accounting of Section 7.4 runs through the components of slack, not the regions of the curve.

FIGURE 4: Common Factor of PLS and the Share of Tight Labor Markets



Notes: The figure plots the estimated common factor of PLS (first principal component of PLS; left axis) against the percentage of sub-sector-country pairs above their estimated tightness threshold (right axis). Thresholds are the random-forest estimates of Appendix D.7.

robustness exercise and a block bootstrap that re-estimates the factor structure on each resample are reported in Appendix D.7.

Three implications follow from Table 11. First, the non-linearity helps explain the difference between the synchronized and idiosyncratic slopes documented in Section 7.4. The two slopes are similar in normal states but diverge when aggregate tightness is high. The larger synchronized coefficient in the baseline linear specification therefore primarily reflects the stronger response estimated during periods of widespread labor shortages.

Second, what steepens the curve is not the difficulty of hiring in any one sub-sector but the synchronized, persistent tightening that moves many sectors and countries at once – consistent with state-dependent pricing, in which the share of firms adjusting increases when cost pressures are substantial, persistent and shared across firms, and with expectations and wage norms that answer aggregate rather than local conditions³⁹.

Third, the result ties the non-linearity to Failure II. The steepening lives in the synchronized component interacted with the aggregate state – the deleted variation times a function of its

³⁹A non-exhaustive list of papers on the topic includes Cavallo et al. (2024); Blanco et al. (2024); Karadi et al. (2023); Gagnon (2009).

own level – so a time-dummy design deletes not only the component that carries the larger slope but the very margin along which the slope changes.

TABLE 11: The non-linearity lives in the synchronized component

	Tight aggregate state D_t defined by	
	$S_t \geq \text{median}$	$S_t \geq 75\text{th percentile}$
Synchronized component (α_c , normal state)	0.016 (0.009) [0.028]	0.009 (0.006) [0.019]
Synchronized $\times$ tight (γ_c)	0.160 (0.016) [0.042] {0.092, 0.262}	0.223 (0.014) [0.038] {0.157, 0.305}
Idiosyncratic component (α_x , normal state)	0.025 (0.007) [0.014]	0.021 (0.004) [0.009]
Idiosyncratic $\times$ tight (γ_x)	-0.006 (0.010) [0.016] {-0.036, 0.024}	-0.041 (0.009) [0.015] {-0.069, -0.012}
Observations	9,094	9,094

Notes: Estimates of equation (15): interactive fixed effects at rank one with sub-sector and country effects, the expectations proxy and the baseline controls; the components $\mu_{ij}F_t^x$ and ξ_{ijt} are the stagewise construction of Section 7.4. S_t is the share of pairs above their random-forest tightness threshold (Figure 4); the cutoffs s^* are the sample median and 75th percentile of S_t , and the main effect of D_t is included. Robust standard errors in parentheses; brackets report the standard deviation across a block bootstrap ($B = 200$); braces report the bootstrap 95 percent percentile intervals for the interaction terms. γ_c is the state-dependence of the projection on the synchronized component.

It should be noted that, the synchronized coefficient retains a projection interpretation in this case. The internal instrument provides insufficient identifying variation to estimate their state dependence precisely, so the non-linear results characterize a conditional relationship rather than a causal effect.

7.6 Robustness Checks and Additional Analyses

We present a range of additional robustness checks to further assess the strength of our main points, complementing those discussed in the main text. We notice that the overall robustness of the results for the application is particularly comforting, given the notorious sensitivity of Phillips curve estimates to different specifications and variable choices.

Additional Monte Carlo Evidence. Appendix C reports a further battery of simulations, of which five results deserve mention. *First*, Experiment G9 confirms that heteroskedasticity-robust standard errors and corrections based on time fail to deliver nominal coverage (e.g., Driscoll-Kraay). Clustering by unit restores coverage when loadings are uncorrelated, but neither eliminates the variance floor nor corrects the bias when loadings are correlated. *Second*, both coefficient bias and coverage distortions increase with loading heterogeneity, as Lemmas 1 and 3 predict. *Third*, the group $\times$ time dummy ladder of Corollary 2 relocates the failures down the hierarchy without removing them, and under measurement error finer dummies move the estimate *further* from the true coefficient even as their bias term shrinks. *Fourth*, concentration is self-revealing: rank-one IFE degrades in the product of feedback strength and the largest unit’s share, but the omitted common component is sufficiently strong for a second factor to absorb exactly where it is strong enough to matter (Experiment G12). *Fifth*, the measurement-error boundary of Experiment G6 holds in the shared-factor design too (Experiment G13).

Mean-Group and Common Correlated Effects Estimators. The natural alternative to IFE for a multifactor error structure is the common correlated effects estimator of Pesaran (2006), which proxies the factors with the cross-sectional averages of the regressor and the outcome; the plain mean-group estimator of Pesaran and Smith (1995) averages unit-by-unit slopes with no augmentation. Experiment G14 runs both against Time FE and IFE. Mean-group does not help: averaging heterogeneous slopes leaves the loading product in every unit’s estimate. CCE has three main limits. It targets the micro-local slope only, since partialling out the averages discards the synchronized component. When only the confounding factor is mean-loaded, the averages proxy the wrong factor and much of the Time FE bias survives. At the boundary of its rank condition coverage collapses as the panel lengthens. When the cross-sectional averages adequately capture the common factors, CCE can perform comparably to IFE. However, this condition is more difficult to assess empirically than the factor rank in IFE, for which established selection criteria are available.

Sample and endogeneity. The results are robust to controlling for the post-pandemic period, to winsorizing, to lagging the components one and two quarters: both component slopes and the conventional time fixed effects estimate remain broadly stable across these specifications (Appendix D.7, Table A.19). Panel unit-root tests reject a pervasive unit root in every series (Table A.27). The stability of both component slopes when PLS is lagged by one or two quarters suggests that the results are not driven solely by contemporaneous feedback from inflation to reported shortages. We further assess dynamic feedback and

anticipatory reporting through two exercises in Appendix D.7. The lead–lag results provide limited evidence of persistent reverse causality: the coefficient on PLS four quarters ahead is small and statistically insignificant. Positive coefficients at shorter leads are consistent with the forward-looking response of inflation to expected future tightness, although they cannot distinguish this mechanism from anticipatory reporting. Conversely, past inflation is only weakly associated with subsequent shortages, and this association becomes statistically insignificant within a year. Taken together, these results support the robustness of the estimated relationship.

Supply Shortages. The main estimation proxies supply chain disruptions with the percentage of firms reporting shortages of material and/or equipment. The BCS offers an alternative, capacity constraints⁴⁰, which have been particularly pronounced in Europe (Lane, 2024). The substitution leaves the pattern intact: the interactive fixed effects columns keep a significantly positive slope while the time FE column sits at zero; on the harmonized sample the component slopes are unchanged by the substitution (Table A.19).

From Tight Labor Markets to Wages. The canonical transmission mechanism from tight labor markets to inflation runs through wages, and PLS captures that channel: Appendix D.7 shows that increases in perceived shortages are associated with significant country-level wage growth, consistent with PLS measuring economically meaningful constraints. The appendix also discusses whether the reverse channel of Afrouzi et al. (2024) – price inflation inducing job search and raising vacancies – affects the interpretation of the measure.

Outliers. Two exercises assess the sensitivity of the results to extreme observations and economically large sub-sector-country pairs. First, winsorizing inflation and PLS at the 1st and 99th percentiles leaves both component slopes largely unchanged (Table A.19). Second, excluding the largest pairs by economic weight and repeating the full two-stage procedure yields an estimated common factor of slack closely correlated with its full-sample counterpart (Table A.21). These findings indicate that the slope estimates are robust to outliers and that the estimated common tightness cycle is not driven by the largest units.

Clustering and standard errors. Inference for the linear specifications remains robust to panel clustering, two-way clustering and panel bootstrap procedures (Tables A.28 and A.29).

⁴⁰Reported as a time series for each country-sub-sector pair, from the question: “*At what capacity is your company currently operating (as a percentage of full capacity)? The company is currently operating at X% of full capacity*”.

For the non-linear specification, a block bootstrap that re-estimates the factor structure in each replication yields confidence intervals excluding zero for both interaction coefficients (Table A.24). The synchronized slope increases above the threshold, while the idiosyncratic slope declines. This pattern also holds when the thresholds are re-estimated using PLS net of its common component (Table A.25).

Difference-in-Differences Extension. The three econometric problems identified above also arise in panel designs with discrete treatments. Online Appendix OA.1.2 extends the framework to staggered difference-in-differences. Correlation between treatment timing and latent outcome loadings can violate parallel trends, while unmodelled factor dependence can distort conventional inference even under random assignment. These problems are distinct from the weighting issues addressed by estimators allowing for heterogeneous treatment effects. Under the maintained identification and rank conditions, IFE accounts for heterogeneous factor exposure and delivers approximately unbiased estimates with near-nominal coverage in the staggered-adoption simulations. An event-study illustration further shows that pre-trend tests need not detect factor confounding when the common shock emerges only after treatment.

8 Conclusion

Cross-sectional Phillips curve estimates depend on how the empirical design treats common variation. When sectors differ in their exposure to aggregate forces, time fixed effects remove the average response to those forces but leave their heterogeneous transmission unaccounted for. This can bias the estimated slope and generate a component of sampling uncertainty that does not vanish as the panel lengthens. Moreover, when inflation responds differently to idiosyncratic and synchronized tightness, a pooled coefficient combines economically distinct responses. The apparent flatness of the Phillips curve can therefore reflect both an identification problem and the composition of the variation used to estimate it.

This paper develops a two-step extension of interactive fixed effects that separates these responses. Defactoring the regressor identifies the micro-local slope without observed exposure measures or an identified aggregate shock. Modelling the outcome’s factor structure also addresses the bias and variance floor associated with heterogeneous exposure. Identifying the synchronized response requires an additional restriction on aggregate variation: the cross-section helps recover the common component and its exposures, but does not by itself establish the causal effect of that component.

The euro area application demonstrates the economic importance of this distinction. Firms' reported labor shortages provide a measure of tightness at the sub-sector-country level. Using this measure, the estimated response to a sector's own tightness is approximately three times the conventional time fixed effects estimate. The response to synchronized tightness is substantially larger still. Under the internal instrument's exclusion restriction, a one-standard-deviation increase in synchronized tightness raises inflation by about 1.3 percentage points, compared with approximately 0.25 percentage points for an idiosyncratic increase.

The non-linear estimates further indicate that the stronger relationship with synchronized tightness is concentrated in tight aggregate states. The idiosyncratic slope exhibits no comparable steepening. This state dependence suggests that the prevalence of shortages is an important dimension of the inflation–tightness relationship.

The methodological implications extend to other applications that infer relationships from disaggregated data. Heterogeneous exposure determines whether common forces are adequately controlled for, while the composition of identifying variation determines which response a coefficient represents. Distinguishing these questions is essential when using cross-sectional evidence to assess aggregate outcomes. In the Phillips curve application, it reveals that the inflationary implications of labor shortages depend on whether those shortages are confined to individual markets or shared across the economy.

References

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program. *Journal of the American Statistical Association* 105(490), 493–505.
- Adão, R., C. Arkolakis, and F. Esposito (2019). Spatial linkages, global shocks, and local labor markets: Theory and evidence. *Cowles Foundation Discussion Paper No. 2163*.
- Adão, R., M. Kolesár, and E. Morales (2019). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics* 134(4), 1949–2010.
- Afrouzi, H., A. Blanco, A. Drenik, and E. Hurst (2024). A theory of how workers keep up with inflation. Technical report, National Bureau of Economic Research.
- Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81(3), 1203–1227.
- Albuquerque, B. and U. Baumann (2017). Will US inflation awake from the dead? The role of slack and non-linearities in the Phillips curve. *Journal of Policy Modeling* 39(2), 247–271.
- Almuzara, M. and V. Sancibrián (2024). Micro responses to macro shocks. *FRB of New York Staff Report* (1090).
- Altavilla, C., L. Brugnolini, R. S. Gürkaynak, R. Motto, and G. Ragusa (2019). Measuring euro area monetary policy. *Journal of Monetary Economics* 108, 162–179.
- Alvarez-Blaser, S., R. Auer, S. M. Lein, and A. A. Levchenko (2025). The granular origins of inflation. Working paper.
- Ampudia, M., M. Lombardi, and T. Renault (2024). *The wage-price pass-through across sectors: evidence from the euro area*. Number 2948. ECB Working Paper.
- Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20(1), 46–63.
- Andrews, D. W. K. (2005). Cross-section regression with common shocks. *Econometrica* 73(5), 1551–1585.

- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021). Synthetic difference-in-differences. *American Economic Review* 111(12), 4088–4118.
- Arkhangelsky, D. and V. Korovkin (2024). On policy evaluation with aggregate time-series instruments. *CEPR Discussion Paper No. 18931*. arXiv:1905.13660.
- Auer, R., M. Pedemonte, and R. Schoenle (2025). Sixty years of global inflation: a post-GFC update. In *Research Handbook on Inflation*, pp. 390–419. Edward Elgar Publishing.
- Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review* 103(6), 2121–2168.
- Babb, N. R. and A. K. Detmeister (2017). Nonlinearities in the Phillips curve for the United States: Evidence using metropolitan data. *Finance and Economics Discussion Series 2017-070*, Board of Governors of the Federal Reserve System.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229–1279.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Ball, L., D. Leigh, and P. Mishra (2022). Understanding US inflation during the COVID-19 era. *Brookings Papers on Economic Activity* 2022(2), 1–80.
- Barnichon, R. and A. H. Shapiro (2024). Phillips meets Beveridge. *Journal of Monetary Economics* 148, 103660.
- Beaudry, P., C. Hou, and F. Portier (2025). On the fragility of the nonlinear Phillips curve view of recent inflation. Technical report, National Bureau of Economic Research.
- Benigno, P. and G. B. Eggertsson (2023). It’s baaack: The surge in inflation in the 2020s and the return of the non-linear Phillips curve. Technical report, National Bureau of Economic Research.
- Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1), 249–275.
- Beschin, A., J. Paredes, G. Polichetti, and T. Renault (2025). The slope of the euro area price Phillips curve: evidence from regional data. *ECB Working Paper No. 3133*.

- Blanco, A., C. Boar, C. Jones, and V. Midrigan (2024). *Non-linear inflation dynamics in menu cost economies*. Number w32094. National Bureau of Economic Research.
- Bonam, D., J. de Haan, and D. Van Limbergen (2021). Time-varying wage Phillips curves in the euro area with a new measure for labor market slack. *Economic Modelling* 96, 157–171.
- Bonthuis, B., V. Jarvis, and J. Vanhala (2016). Shifts in euro area Beveridge curves and their determinants. *IZA Journal of Labor Policy* 5(1), 20.
- Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies* 89(1), 181–213.
- Callaway, B. and S. Karami (2023). Treatment effects in interactive fixed effects models with a small number of time periods. *Journal of Econometrics* 233(1), 184–208.
- Callaway, B. and P. H. Sant’Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics* 225(2), 200–230.
- Canova, F. (2024). Should we trust cross-sectional multiplier estimates? *Journal of Applied Econometrics* 39(4), 589–606.
- Cavallo, A., F. Lippi, and K. Miyahara (2024). Large shocks travel fast. *American Economic Review: Insights* 6(4), 558–574.
- Chaudhary, M., J. Z. Fu, and H. Zhou (2026). Anatomy of the treasury market: Who moves yields? FMG Discussion Paper No. 959, London School of Economics.
- Chodorow-Reich, G. (2020). Regional data in macroeconomics: Some advice for practitioners. *Journal of Economic Dynamics and Control* 115, 103875.
- Chodorow-Reich, G., P. T. Nenov, and A. Simsek (2021). Stock market wealth and the real economy: A local labor market approach. *American Economic Review* 111(5), 1613–1657.
- Ciccarelli, M. and B. Mojon (2010). Global inflation. *The Review of Economics and Statistics* 92(3), 524–535.
- Daniele, M. (2024). Selecting the number of factors in approximate factor models using group variable regularization. *Econometric Reviews* 43(10), 796–823.
- de Chaisemartin, C. and X. D’Haultfoeulle (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110(9), 2964–2996.

- Ditzen, J. and S. Reese (2023). xtnumfac: A battery of estimators for the number of common factors in time series and panel-data models. *The Stata Journal* 23(2), 438–454.
- Donaldson, P. (2025). Cross-sectional identification with heterogeneous exposure to general equilibrium effects. Job market paper, University of California San Diego.
- Eser, F., P. Karadi, P. R. Lane, L. Moretti, and C. Osbat (2020). The Phillips curve at the ECB. *The Manchester School* 88(S1), 50–85.
- Faber, M. and G. Züllig (2026). Kinky Europe: Evidence from the regional Phillips curve in the euro area. *European Economic Review* 189, 105439.
- Fitzgerald, T., C. Jones, M. Kulish, and J. P. Nicolini (2024). Is there a stable relationship between unemployment and future inflation? *American Economic Journal: Macroeconomics* 16(4), 114–142.
- Florio, A., D. Siena, and R. Zago (2025). Global value chains and the Phillips curve: A challenge for monetary policy. *European Economic Review* 174, 104966.
- Furlanetto, F. and A. Lepetit (2025). The slope of the Phillips curve. In *Research Handbook on Inflation*, pp. 167–185. Edward Elgar Publishing.
- Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica* 79(3), 733–772.
- Gabaix, X. and R. S. J. Koijen (2024). Granular instrumental variables. *Journal of Political Economy* 132(7), 2274–2303.
- Gagliardini, P., E. Ossola, and O. Scaillet (2019). A diagnostic criterion for approximate factor structure. *Journal of Econometrics* 212(2), 503–521.
- Gagliardone, L., M. Gertler, S. Lenzu, and J. Tielens (2025). Anatomy of the Phillips curve: micro evidence and macro implications. *American Economic Review* 115(11), 3941–3974.
- Gagnon, E. (2009). Price setting during low and high inflation: Evidence from Mexico. *The Quarterly Journal of Economics* 124(3), 1221–1263.
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the New Keynesian framework and its applications*. Princeton University Press.
- Gitti, G. (2024). Nonlinearities in the regional Phillips curve with labor market tightness. Unpublished manuscript, Brown University.

- Gobillon, L. and T. Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics* 98(3), 535–551.
- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review* 110(8), 2586–2624.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225(2), 254–277.
- Hazell, J., J. Herreño, E. Nakamura, and J. Steinsson (2022). The slope of the Phillips curve: evidence from US states. *The Quarterly Journal of Economics* 137(3), 1299–1344.
- Hong, G. H., Z. Kóczán, W. Lian, and M. S. Nabar (2018). *More slack than meets the eye? Recent wage dynamics in advanced economies*. International Monetary Fund.
- Hooper, P., F. S. Mishkin, and A. Sufi (2020). Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating? *Research in Economics* 74(1), 26–62.
- Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115(1), 53–74.
- Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises: The role of information shocks. *American Economic Journal: Macroeconomics* 12(2), 1–43.
- Karadi, P., J. Amann, J. S. Bachiller, P. Seiler, and J. Wursten (2023). Price setting on the two sides of the Atlantic – Evidence from supermarket scanner data. *Journal of Monetary Economics* 140, S1–S17.
- Kiss, A., A. Turrini, and A. Vandeplas (2022). II. Slack vs tightness in euro area labour markets: growing mismatch after COVID-19? *Quarterly Report on the Euro Area* 21(2), 19–28.
- Lane, P. R. (2024). The 2021-2022 inflation surges and the monetary policy response through the lens of macroeconomic models. *SUERF Policy Note* 364.
- Levy, A. (2019). The euro area Phillips curve: identification with regional data. *Unpublished manuscript, Department of Economics, Massachusetts Institute of Technology*.
- Majerovitz, J. and K. A. Sastry (2023). How much should we trust regional-exposure designs? Working Paper 2023-018, Federal Reserve Bank of St. Louis.

- Manuel, E. (2026). Missing intercept, biased slope: Identification in general equilibrium. Working paper, London School of Economics. Version of 14 September 2026.
- Matthes, C., N. Nagasaka, and F. F. Schwartzman (2025). Estimating the missing intercept. *FRB Richmond Working Paper No. 25-12*.
- Mavroeidis, S., M. Plagborg-Møller, and J. H. Stock (2014). Empirical evidence on inflation expectations in the New Keynesian Phillips curve. *Journal of Economic Literature* 52(1), 124–188.
- McKay, A. and C. K. Wolf (2023). What can time-series regressions tell us about policy counterfactuals? *Econometrica* 91(5), 1695–1725.
- McLeay, M. and S. Tenreyro (2020). Optimal inflation and the identification of the Phillips curve. *NBER Macroeconomics Annual* 34(1), 199–255.
- Moll, B. and O. Hanney (2025). The ‘missing intercept’ problem with going from micro to macro. VoxDev.
- Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics* 32(3), 385–397.
- Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2), 334–338.
- Murphy, A. (2017). Is the US Phillips curve convex? Some metro level evidence. *Manuscript*.
- Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review* 104(3), 753–792.
- Nakamura, E. and J. Steinsson (2018). Identification in macroeconomics. *Journal of Economic Perspectives* 32(3), 59–86.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92(4), 1004–1016.
- Ottobello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica* 88(6), 2473–2502.
- Pennings, S. (2021). Cross-region transfer multipliers in a monetary union: Evidence from social security and stimulus payments. *American Economic Review* 111(5), 1689–1719.

- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967–1012.
- Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68(1), 79–113.
- Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica* 25(100), 283–299.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT Press.
- Reis, R. (2026). Why did inflation rise and fall in 2021–2024? Channels and evidence from expectations. *Annual Review of Economics* 18, 473–503.
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights* 4(3), 305–322.
- Sarto, A. P. (2025). Recovering macro elasticities from regional data. Working paper, University of Illinois, December 2025.
- Słoczyński, T. (2022). Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights. *Review of Economics and Statistics* 104(3), 501–509.
- Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2), 175–199.
- Wolf, C. K. (2023). The missing intercept: A demand equivalence approach. *American Economic Review* 113(8), 2232–2269.
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1), 57–76.

Appendix

A Formal Assumptions	57
B Additional Proofs and Technical Discussion	58
B.1 Proof of Lemma 1	58
B.2 Proof of Lemma 2	60
B.3 Proof of Lemma 3	61
B.4 What Variance Estimators Can and Cannot Price	66
B.5 Distinct Aggregates: The Two-Factor Bias and Signal Deletion	67
B.6 Proof of Proposition 1	70
C Complements to the Method	75
D Complements to the Application	92
D.1 Detailed Data Sources and Descriptions	92
D.2 List of Manufacturing Sub-Sectors	95
D.3 Summary Statistics	96
D.4 Estimates of the Common Component (Rank Selection)	96
D.5 Estimated Common Factor and Distribution of Loadings	97
D.6 Complements to the Estimand Decomposition	103
D.7 Complements to the Robustness Checks	104
D.8 Im, Pesaran and Shin (2002) Unit Root Test	114
D.9 Alternative Standard Errors	114

A Formal Assumptions

[↔ Back to reference in main text](#)

This appendix states formally the four regularity conditions summarized in Section 4. They are standard in the factor-model literature; the granularity condition specific to this paper, Assumption 1, is stated in the main text.

Assumption 2 (Loading heterogeneity). $\sigma_\lambda^2 > 0$ and $\sigma_\mu^2 > 0$: factor loadings vary genuinely across units.

Assumption 3 (Factor and loading strength). (i) F_t is strictly stationary with $\mathbb{E}[F_t^4] \leq M$ and $T^{-1} \sum_{t=1}^T F_t^2 \xrightarrow{p} \Sigma_F > 0$; (ii) $\mathbb{E}[\lambda_i^4] \leq M$, $\mathbb{E}[\mu_i^4] \leq M$, and $N^{-1} \sum_i \lambda_i^2 \xrightarrow{p} \Sigma_\lambda > 0$, $N^{-1} \sum_i \mu_i^2 \xrightarrow{p} \Sigma_\mu > 0$. The factor is strong in the sense of Bai (2009, Assumption B); under the loading distribution of Section 3, $\Sigma_\lambda = \bar{\lambda}^2 + \sigma_\lambda^2$ and $\Sigma_\mu = \bar{\mu}^2 + \sigma_\mu^2$, so Assumption 2 implies the strict inequalities.

Assumption 4 (Rank condition). The regressor retains idiosyncratic variation once the factor space and the fixed effects are projected out: $\inf_F D(F) > 0$, where $D(F)$ is the second-moment matrix of the regressor after projecting out the candidate factor process and cross-sectional averages. In the one-factor DGP (2) this holds if and only if $\text{Var}(\xi_{it}) > 0$: the idiosyncratic component ξ_{it} is what survives the projection⁴¹. The assumption also requires the factor not to be collinear with the unit effects, which the AR(1) specification of F_t with $\text{Var}(F_t) > 0$ guarantees.

Assumption 5 (Weak error dependence). The cross-sectional covariance matrix of the errors ε_{it} has eigenvalues bounded as $N \rightarrow \infty$, their autocovariances are absolutely summable in t (the weak-dependence condition of Appendix B.3), and eighth moments are bounded. Together with idiosyncratic exogeneity ($\mathbb{E}[\xi_{is}\varepsilon_{it}] = 0$ for all s, t , Section 3) and the independence of ε_{it} from the loadings and $\{F_t\}$: the error is exogenous to the regressors, loadings and factors at all leads and lags.

⁴¹Notice that when the regressor's common component loads on an aggregate distinct from the outcome's error factor, that component supplies further identifying variation – the signal that IFE retains and time dummies delete – but only under the weaker requirement that D be positive at the true factor process (Bai 2009). Appendix B.5 develops this two-factor case.

B Additional Proofs and Technical Discussion

B.1 Proof of Lemma 1

[↔ Back to reference in main text](#)

Lemma 1 (Time FE Bias). *Under the DGP (1)–(2), with (λ_i, μ_i) i.i.d. across units with variance σ^2 , and correlation $\rho_{\lambda\mu}$; F_t strictly stationary with $\text{Var}(F_t) > 0$; ξ_{it} and ε_{it} mean-zero, independent of all loadings and of $\{F_t\}$; $\mathbb{E}[\xi_{is}\varepsilon_{it}] = 0$ for all s, t ; and $\text{Var}(\tilde{x}_{it}) > 0$:*

$$\text{plim}_{N,T \rightarrow \infty} (\hat{\alpha}_{\text{TimeFE}} - \alpha) = \frac{\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) \cdot \text{Var}(\tilde{F}_t)}{\text{Var}(\tilde{x}_{it})}, \quad (16)$$

where tildes denote residuals after projecting out unit and time means. Under i.i.d. loadings, $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = \rho_{\lambda\mu}\sigma^2(1 - N^{-1})$, so under idiosyncratic exogeneity the bias equals zero if and only if $\rho_{\lambda\mu}\sigma^2 = 0$.

Proof. Step 1: Within transformation. In a balanced panel the two-way within estimator applies the Frisch–Waugh projection $M_{[D_i, D_t]}$, which in a balanced panel equals the double-demeaning operator

$$\tilde{z}_{it} := z_{it} - \bar{z}_{i\cdot} - \bar{z}_{\cdot t} + \bar{z},$$

where $\bar{z}_{i\cdot} = T^{-1} \sum_t z_{it}$, $\bar{z}_{\cdot t} = N^{-1} \sum_i z_{it}$, and $\bar{z} = (NT)^{-1} \sum_{i,t} z_{it}$.

Step 2: Factor terms under the within transformation. For $x_{it} = \mu_i F_t + \xi_{it}$:

$$\bar{x}_{i\cdot} = \mu_i \bar{F} + \bar{\xi}_{i\cdot}, \quad \bar{x}_{\cdot t} = \bar{\mu} F_t + \bar{\xi}_{\cdot t}, \quad \bar{x} = \bar{\mu} \bar{F} + \bar{\xi},$$

so

$$\tilde{x}_{it} = (\mu_i - \bar{\mu})(F_t - \bar{F}) + \tilde{\xi}_{it} =: \tilde{\mu}_i \tilde{F}_t + \tilde{\xi}_{it}.$$

Identically, $\tilde{y}_{it} = \alpha \tilde{x}_{it} + \tilde{\lambda}_i \tilde{F}_t + \tilde{\varepsilon}_{it}$, where $\tilde{\lambda}_i := \lambda_i - \bar{\lambda}$. The unit and time fixed effects γ_i and δ_t vanish exactly under double-demeaning.

Step 3: OVB decomposition. The within estimator satisfies

$$\hat{\alpha}_{\text{TimeFE}} - \alpha = \frac{(NT)^{-1} \sum_{i,t} \tilde{x}_{it} (\tilde{\lambda}_i \tilde{F}_t + \tilde{\varepsilon}_{it})}{(NT)^{-1} \sum_{i,t} \tilde{x}_{it}^2}.$$

Because the within projection is idempotent, $\sum_{i,t} \tilde{x}_{it} \tilde{\varepsilon}_{it} = \sum_{i,t} \tilde{x}_{it} \varepsilon_{it}$. Substituting $\tilde{x}_{it} =$

$\tilde{\mu}_i \tilde{F}_t + \tilde{\xi}_{it}$ splits this term in two. The first piece, $(NT)^{-1} \sum_{i,t} \tilde{\mu}_i \tilde{F}_t \varepsilon_{it}$, has probability limit zero because ε_{it} is mean-zero and independent of the loadings and of $\{F_t\}$. The second piece, $(NT)^{-1} \sum_{i,t} \tilde{\xi}_{it} \varepsilon_{it}$, involves the covariances $\mathbb{E}[\xi_{is} \varepsilon_{it}]$ at all leads and lags, since the demeaned $\tilde{\xi}_{it}$ contains the unit and time means of the ξ draws; under idiosyncratic exogeneity ($\mathbb{E}[\xi_{is} \varepsilon_{it}] = 0$ for all s, t) every such covariance vanishes, and the piece converges to zero in probability by a law of large numbers. The convergence of $T^{-1} \sum_t \tilde{F}_t^2$ and of the mixed idiosyncratic products draws on the weak-dependence structure of the data-generating process – the AR(1) specifications for F_t and ξ_{it} in Section 3, which deliver absolutely summable autocovariances – together with finite fourth moments for $(\xi_{it}, \varepsilon_{it}, F_t)$. These are the conditions under which the ergodic theorem applies term by term; strict stationarity alone would not suffice. Without idiosyncratic exogeneity the estimator picks up the additional term $\mathbb{E}[\tilde{\xi}_{it} \tilde{\varepsilon}_{it}] / \text{Var}(\tilde{x}_{it})$ and the zero-bias characterisation below fails; the condition is not implied by the independence of ε_{it} from the loadings and the factor. Hence

$$\text{plim}(\hat{\alpha}_{\text{TimeFE}} - \alpha) = \frac{\text{plim} (NT)^{-1} \sum_{i,t} \tilde{x}_{it} \tilde{\lambda}_i \tilde{F}_t}{\text{Var}(\tilde{x}_{it})}.$$

Step 4: Numerator. Substituting $\tilde{x}_{it} = \tilde{\mu}_i \tilde{F}_t + \tilde{\xi}_{it}$:

$$\frac{1}{NT} \sum_{i,t} \tilde{x}_{it} \tilde{\lambda}_i \tilde{F}_t = \frac{1}{N} \sum_i \tilde{\mu}_i \tilde{\lambda}_i \cdot \frac{1}{T} \sum_t \tilde{F}_t^2 + \frac{1}{NT} \sum_{i,t} \tilde{\xi}_{it} \tilde{\lambda}_i \tilde{F}_t.$$

The second term converges in probability to zero because ξ_{it} is independent of λ_i and F_t (so $\mathbb{E}[\tilde{\xi}_{it} \tilde{\lambda}_i \tilde{F}_t] = 0$) and satisfies a law of large numbers under standard moment assumptions. By the LLN applied separately in the cross-section and in the time dimension, the first term converges to $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) \cdot \text{Var}(\tilde{F}_t)$, establishing (3).

Step 5: Loading covariance under i.i.d. draws. Since $\tilde{\mu}_i = \mu_i - \bar{\mu}$ and $\tilde{\lambda}_i = \lambda_i - \bar{\lambda}$,

$$\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = \text{Cov}(\mu_i, \lambda_i) - \text{Cov}(\mu_i, \bar{\lambda}) - \text{Cov}(\bar{\mu}, \lambda_i) + \text{Cov}(\bar{\mu}, \bar{\lambda}).$$

Under i.i.d. loadings, $\text{Cov}(\mu_i, \bar{\lambda}) = \text{Cov}(\mu_i, N^{-1} \sum_j \lambda_j) = N^{-1} \rho_{\lambda\mu} \sigma^2$, and similarly for the other cross terms, giving

$$\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = \rho_{\lambda\mu} \sigma^2 \left(1 - \frac{1}{N} \right).$$

Step 6: Necessary and sufficient assumptions for zero bias. Under idiosyncratic exogeneity the estimation error consists of the factor term alone, so the right-hand side of

(3) equals zero if and only if $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = 0$, which under i.i.d. loadings requires $\rho_{\lambda\mu}\sigma^2 = 0$. □ □

B.2 Proof of Lemma 2

[↔ Back to reference in main text](#)

The lemma's first claim is immediate: under (4), conditioning on both factors, $\partial \mathbb{E}[y_{it} \mid x_{it}, F_t^x, F_t^y] / \partial x_{it} = \alpha_x$, since the synchronized component $\mu_i F_t^x$ is fixed by the conditioning and only ξ_{it} moves with x_{it} . For the composition identity, write the synchronized component as $c_{it} \equiv \mu_i F_t^x$, so that, using $x_{it} = c_{it} + \xi_{it}$, the outcome equation (4) reads $y_{it} = \alpha_x x_{it} + (\alpha_c - \alpha_x) c_{it} + \lambda_i F_t^y + \gamma_i + \varepsilon_{it}$. Fix an estimator e that absorbs a set of effects and then fits a single linear slope, and let $\tilde{x}_{it}^{(e)}$ be the residual of x_{it} after that absorption. By Frisch–Waugh–Lovell the estimator's probability limit is

$$\text{plim } \hat{\alpha}_e = \frac{\text{Cov}(\tilde{x}_{it}^{(e)}, y_{it})}{\text{Var}(\tilde{x}_{it}^{(e)})}.$$

Three of the terms in y_{it} drop out. The absorbed effects are orthogonal to $\tilde{x}_{it}^{(e)}$ by construction; the residual factor structure is orthogonal to it by the zero-bias condition maintained in the statement; and ε_{it} is orthogonal to it by idiosyncratic exogeneity. For the first surviving term, note that $x_{it} = \tilde{x}_{it}^{(e)} + a_{it}^{(e)}$ where $a_{it}^{(e)}$ collects the absorbed part, so $\text{Cov}(\tilde{x}_{it}^{(e)}, x_{it}) = \text{Var}(\tilde{x}_{it}^{(e)})$ and that term contributes exactly α_x . The second surviving term contributes $(\alpha_c - \alpha_x) \text{Cov}(\tilde{x}_{it}^{(e)}, c_{it}) / \text{Var}(\tilde{x}_{it}^{(e)}) = (\alpha_c - \alpha_x) w_e$ by the definition (5). Collecting the two gives (6). Applying it to $e = \text{FA}$ and $e = \text{TFE}$ and differencing gives (7), the common term α_x canceling. □

Two observations follow from the algebra. The result is an identity given the component model and the zero-bias condition, not an approximation: whatever the two estimators do to the data, the gap they produce is $(\alpha_c - \alpha_x)$ times the difference in their synchronized-component weights. And at $\alpha_c = \alpha_x$ the identity delivers $\text{plim } \hat{\alpha}_e = \alpha_c = \alpha_x$ for every e , so no absorption scheme can move the estimand – the common-response case in which time demeaning costs precision alone. The argument is generic in the absorption scheme and in the panel's index structure: in the application it is applied in the two-way (ijt) notation of Section 7.4, with the controls and additive effects of (14) among the absorbed terms.

B.3 Proof of Lemma 3

↔ *Back to reference in main text*

Lemma 3 (Variance Floor: Convergence at $\sqrt{N}$). *Consider the DGP (1)–(2) with the same assumptions as Lemma 1, loadings of finite fourth moment, the weak-dependence conditions imposed in Step 1 below, and $\rho_{\lambda\mu} = 0$. Then, as $N, T \rightarrow \infty$,*

$$\sqrt{N} (\hat{\alpha}_{\text{TimeFE}} - \alpha) \xrightarrow{d} \mathcal{N}(0, V), \quad V = \frac{\kappa [\text{Var}(\tilde{F}_t)]^2}{[\text{Var}(\tilde{x}_{it})]^2}, \quad \kappa \equiv \text{Var}(\tilde{\lambda}_i \tilde{\mu}_i),$$

with $\kappa = \sigma^4 > 0$ under independent loadings. The estimator's sampling variance therefore carries a floor V/N that is invariant to T .

Corollary (Inference Collapse). *Under the conditions of the lemma, $NT \cdot \widehat{\text{Var}}(\hat{\alpha}_{\text{TimeFE}}) \xrightarrow{p} c$ for a finite $c > 0$, so whenever $\sigma^2 > 0$*

$$r_T \equiv \frac{\widehat{\text{se}}(\hat{\alpha}_{\text{TimeFE}})}{\text{se}(\hat{\alpha}_{\text{TimeFE}})} = O_p(T^{-1/2}),$$

and the coverage of nominal 95% confidence intervals, $2\Phi(1.96 r_T) - 1 + o(1)$, tends to zero as T grows.

Proof. Step 1: Decomposition. From the within transformation of Step 2 of the proof of Lemma 1, $\tilde{y}_{it} = \alpha \tilde{x}_{it} + \tilde{\lambda}_i \tilde{F}_t + \tilde{\varepsilon}_{it}$. Substituting into the score of the within estimator gives

$$\hat{\alpha}_{\text{TimeFE}} - \alpha = \frac{\sum_{i,t} \tilde{x}_{it} (\tilde{\lambda}_i \tilde{F}_t + \tilde{\varepsilon}_{it})}{\sum_{i,t} \tilde{x}_{it}^2} =: \frac{A + B}{D},$$

where the three objects carried through the rest of the proof are the factor term, the idiosyncratic term and the denominator,

$$A \equiv \sum_{i,t} \tilde{x}_{it} \tilde{\lambda}_i \tilde{F}_t, \quad B \equiv \sum_{i,t} \tilde{x}_{it} \tilde{\varepsilon}_{it}, \quad D \equiv \sum_{i,t} \tilde{x}_{it}^2.$$

Because the within projection is idempotent, $B = \sum_{i,t} \tilde{x}_{it} \tilde{\varepsilon}_{it}$ as well, and Step 4 works with that form. Throughout the proof we impose the following weak-dependence condition, consistent with the DGP of Section 3: the processes $\{\xi_{it}\}_t$ and $\{\varepsilon_{it}\}_t$ are independent across units, strictly stationary in t , with finite fourth moments and absolutely summable autocovariances, $\sum_{\tau=-\infty}^{\infty} |\gamma_{\xi}(\tau)| < \infty$ and $\sum_{\tau} |\gamma_{\varepsilon}(\tau)| < \infty$; equivalently, each admits a linear $\text{MA}(\infty)$

representation with absolutely summable coefficients. The stationary AR(1) specification $\xi_{it} = \rho_\xi \xi_{i,t-1} + v_{it}$ with $|\rho_\xi| < 1$ maintained in the model satisfies this with $\gamma_\xi(\tau) = \rho_\xi^{|\tau|} \sigma_\xi^2$; the i.i.d. case is the special case $\gamma(\tau) = 0$ for $\tau \neq 0$. The same summability is imposed on $\{F_t\}$ (satisfied by its stationary AR(1) specification), together with $\sum_\tau |\text{Cov}(F_0^2, F_\tau^2)| < \infty$, which holds for a Gaussian AR(1).

Step 2: Mean zero under $\rho_{\lambda\mu} = 0$. Write $\tilde{x}_{it} = \tilde{\mu}_i \tilde{F}_t + \tilde{\xi}_{it}$. Then

$$\mathbb{E}[A] = \mathbb{E} \left[\sum_{i,t} (\tilde{\mu}_i \tilde{F}_t + \tilde{\xi}_{it}) \tilde{\lambda}_i \tilde{F}_t \right] = \sum_{i,t} \mathbb{E}[\tilde{\mu}_i \tilde{\lambda}_i] \mathbb{E}[\tilde{F}_t^2] + \sum_{i,t} \mathbb{E}[\tilde{\xi}_{it}] \mathbb{E}[\tilde{\lambda}_i \tilde{F}_t] = N \mathbb{E}[\tilde{\mu}_i \tilde{\lambda}_i] \bar{S}_F,$$

where $\bar{S}_F \equiv \sum_{t=1}^T \mathbb{E}[\tilde{F}_t^2] = O(T)$, the first factorization uses the independence of the loadings from $\{F_t\}$, and the second term vanishes because the ξ -process is independent of the loadings and of $\{F_t\}$ with $\mathbb{E}[\tilde{\xi}_{it}] = 0$. Under $\rho_{\lambda\mu} = 0$, $\mathbb{E}[\tilde{\mu}_i \tilde{\lambda}_i] = \text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = \rho_{\lambda\mu} \sigma^2 (1 - N^{-1}) = 0$, hence $\mathbb{E}[A] = 0$: the factor of T scales the display but not the conclusion, and the variance rates of Steps 3–5, which carry the time dimension explicitly, are unaffected. $\mathbb{E}[B] = 0$ follows from idiosyncratic exogeneity ($\mathbb{E}[\xi_{is} \varepsilon_{it}] = 0$ for all s, t) together with the independence of ε_{it} from the loadings and $\{F_t\}$.

Step 3: Variance of the factor term A . Decompose $A = A_1 + A_2$ where

$$A_1 \equiv \left(\sum_i \tilde{\mu}_i \tilde{\lambda}_i \right) S_F, \quad S_F \equiv \sum_t \tilde{F}_t^2, \quad A_2 \equiv \sum_{i,t} \tilde{\xi}_{it} \tilde{\lambda}_i \tilde{F}_t.$$

Conditioning on $\{F_t\}$ and using the independence of the ξ -process from λ_i and $\{F_t\}$, one obtains $\text{Cov}(A_1, A_2) = 0$.

Variance of A_1 . The underlying draws (μ_i, λ_i) are i.i.d. across units, but the demeaned pairs $(\tilde{\mu}_i, \tilde{\lambda}_i)$ are not: subtracting the realized cross-sectional means induces a dependence across i , so the pairs are exchangeable rather than independent and the off-diagonal covariances cannot simply be set to zero. They are, however, of smaller order. Write $u_i \equiv \mu_i - \mathbb{E}\mu_i$ and $v_i \equiv \lambda_i - \mathbb{E}\lambda_i$, so that $\sum_i \tilde{\mu}_i \tilde{\lambda}_i = \sum_i u_i v_i - N \bar{u} \bar{v}$, and define

$$\kappa \equiv \text{Var}(u_i v_i),$$

finite by the fourth-moment bound on the loadings and strictly positive whenever the product is non-degenerate. Under $\rho_{\lambda\mu} = 0$ both pieces are mean-zero, and with independent loadings ($\kappa = \sigma^4$) a direct computation gives $\text{Var}(\sum_i u_i v_i) = N\kappa$, $\text{Var}(N \bar{u} \bar{v}) = \kappa$ and

$\text{Cov}(\sum_i u_i v_i, N\bar{u}\bar{v}) = \kappa$, so the demeaning costs exactly one unit of κ :

$$\text{Var}\left(\sum_i \tilde{\mu}_i \tilde{\lambda}_i\right) = (N-1)\kappa = N\kappa(1 + O(N^{-1})).$$

Without independence the same calculation bounds the correction at $O(1)$, so the display holds either way, and the rate the argument uses is unaffected. Under the weak-dependence condition, $\mathbb{E}[S_F] = O(T)$ and $\text{Var}(S_F) \leq T \sum_\tau |\text{Cov}(\tilde{F}_0^2, \tilde{F}_\tau^2)| = O(T)$, so $\mathbb{E}[S_F^2] = O(T^2)$ and $T^{-2}\mathbb{E}[S_F^2] \rightarrow c_F^2$ with $c_F \equiv \text{Var}(\tilde{F}_t)$. Because the loadings are independent of $\{F_t\}$, the law of total variance then gives $\text{Var}(A_1) = N\kappa \mathbb{E}[S_F^2] (1 + O(N^{-1})) = O(NT^2)$.

Variance of A_2 . Conditional on $\{F_t\}$ and the loadings, the unit-level sums $\sum_t \tilde{\xi}_{it} \tilde{F}_t$ are independent across i with mean zero, so

$$\text{Var}(A_2) = \mathbb{E}\left[\sum_i \tilde{\lambda}_i^2 \text{Var}\left(\sum_t \tilde{\xi}_{it} \tilde{F}_t \mid \{F_t\}\right)\right], \quad \text{Var}\left(\sum_t \tilde{\xi}_{it} \tilde{F}_t \mid \{F_t\}\right) = \sum_{s,t} \tilde{F}_s \tilde{F}_t \gamma_\xi(s, t).$$

Absolute summability of the autocovariances of ξ (inherited by $\tilde{\xi}$ up to $O(T^{-1})$ demeaning terms) bounds the inner double sum by $(\sum_\tau |\gamma_\xi(\tau)|) \sum_t \tilde{F}_t^2 + o_p(T) = O_p(T)$, so $\text{Var}(A_2) = N \mathbb{E}[\tilde{\lambda}_i^2] \cdot O(T) = O(NT)$: serial correlation in ξ rescales the constant by the long-run rather than the contemporaneous variance but leaves the rate unchanged.

Since A_1 dominates, $\text{Var}(A) = \text{Var}(A_1) + \text{Var}(A_2) = O(NT^2)$.

Step 4: Variance of the idiosyncratic term B . By idempotence of the within projection, $B = \sum_i \sum_t \tilde{x}_{it} \varepsilon_{it}$, and the unit-level sums are independent across i . Within a unit, the summands are serially dependent through both $\tilde{x}_{it}$ and, if ε is serially correlated, through ε_{it} ; under the weak-dependence condition the joint autocovariances are absolutely summable, so

$$\text{Var}\left(\sum_t \tilde{x}_{it} \varepsilon_{it}\right) \leq T \sum_\tau |\text{Cov}(\tilde{x}_{i0} \varepsilon_{i0}, \tilde{x}_{i\tau} \varepsilon_{i\tau})| = O(T),$$

the long-run variance of the score replacing the contemporaneous second moment. Hence $\text{Var}(B) = O(NT)$, as in the i.i.d. case.

Step 5: Rates relative to D^2 . The denominator is random, so it must be replaced by a limit rather than treated as a normalizing constant; writing $\text{Var}((A+B)/D)$ as $(\text{Var}(A) + \text{Var}(B))/D^2$ would not be a valid step. Under the maintained stationarity and

weak-dependence conditions a law of large numbers gives

$$\frac{D}{NT} = \frac{1}{NT} \sum_{i,t} \tilde{x}_{it}^2 \xrightarrow{p} c_D \equiv \text{Var}(\tilde{x}_{it}) > 0,$$

the positivity being Assumption 4. Hence $D = c_D NT (1 + o_p(1))$ and

$$\hat{\alpha}_{\text{TimeFE}} - \alpha = \frac{A + B}{c_D NT} (1 + o_p(1)).$$

The leading term is a ratio with a deterministic denominator, so the rates transfer directly:

$$\frac{\text{Var}(A)}{(c_D NT)^2} = \frac{O(NT^2)}{O(N^2 T^2)} = O\left(\frac{1}{N}\right), \quad \frac{\text{Var}(B)}{(c_D NT)^2} = \frac{O(NT)}{O(N^2 T^2)} = O\left(\frac{1}{NT}\right),$$

and $\text{Cov}(A, B) = 0$ by Step 3, so no cross term appears. Because $\sqrt{N}(A + B)/(NT)$ is asymptotically tight, the $o_p(1)$ factor perturbs the asymptotic variance by a term of smaller order and the two displays above give the variance of the limiting distribution. The total is thus

$$\text{Var}(\hat{\alpha}_{\text{TimeFE}}) \asymp \frac{C_1}{N}, \quad C_1 = \frac{\kappa \mathbb{E}[S_F^2]}{c_D^2 T^2} + O\left(\frac{1}{T}\right),$$

with $C_1 > 0$ whenever $\sigma^2 > 0$, and this floor is invariant to T . By Step 3, $C_1 \rightarrow \kappa c_F^2 / c_D^2$, which Step 6 identifies as the asymptotic variance V .

Step 6: Asymptotic normality at rate $\sqrt{N}$. Steps 3–5 leave $A_1/(c_D NT)$ as the only term surviving the $\sqrt{N}$ scaling: $\text{Var}(A_2)$ and $\text{Var}(B)$ are both $O(NT)$, so $\sqrt{N}A_2/(c_D NT)$ and $\sqrt{N}B/(c_D NT)$ are $O_p(T^{-1/2})$ and vanish as $T \rightarrow \infty$. Factor the surviving term into its two averages,

$$\frac{A_1}{c_D NT} = \frac{1}{c_D} \left(\frac{1}{N} \sum_i \tilde{\mu}_i \tilde{\lambda}_i \right) \left(\frac{1}{T} \sum_t \tilde{F}_t^2 \right),$$

the second of which converges in probability to $c_F > 0$ by stationarity and summability. For the first, the decomposition of Step 3 gives $N^{-1} \sum_i \tilde{\mu}_i \tilde{\lambda}_i = N^{-1} \sum_i u_i v_i - \bar{u} \bar{v}$, whose second term is $O_p(N^{-1})$ and therefore negligible at the $\sqrt{N}$ scale. The first term is an average of i.i.d. random variables, mean-zero under $\rho_{\lambda\mu} = 0$ and of finite variance κ by the fourth-moment bound, so the Lindeberg–Lévy theorem applies:

$$\sqrt{N} \frac{1}{N} \sum_i \tilde{\mu}_i \tilde{\lambda}_i \xrightarrow{d} \mathcal{N}(0, \kappa).$$

Slutsky's theorem, together with the $(1 + o_p(1))$ factor of Step 5, then delivers

$$\sqrt{N}(\hat{\alpha}_{\text{TimeFE}} - \alpha) \xrightarrow{d} \mathcal{N}(0, V), \quad V = \frac{\kappa c_F^2}{c_D^2} = \frac{\kappa [\text{Var}(\tilde{F}_t)]^2}{[\text{Var}(\tilde{x}_{it})]^2},$$

which is the constant of Step 5. Two features of the rate are worth recording. The scaling is $\sqrt{N}$ and not $\sqrt{NT}$, because the binding randomness is the single cross-sectional draw of loading products, which T does not replicate; and $T \rightarrow \infty$ enters only to discard the idiosyncratic terms, so the floor V/N is what remains however long the panel runs.

Step 7: What the robust estimator recovers. The heteroskedasticity-robust sandwich estimator is

$$\widehat{\text{Var}}(\hat{\alpha}_{\text{TimeFE}}) = D^{-2} \sum_{i,t} \tilde{x}_{it}^2 \hat{u}_{it}^2,$$

where $\hat{u}_{it}$ are within residuals. It treats all NT score contributions as independent, so in probability it converges to $D^{-2} \sum_{i,t} \mathbb{E}[\tilde{x}_{it}^2 u_{it}^2]$, a sum of NT bounded terms each of order $O(1)$. Hence

$$\widehat{\text{Var}}(\hat{\alpha}_{\text{TimeFE}}) = O_p\left(\frac{NT}{N^2 T^2}\right) = O_p\left(\frac{1}{NT}\right).$$

The key reason it misses the floor is that the N per-period contributions to A_1 are perfectly synchronised through the common draw $\tilde{F}_t$: their covariances across units are non-zero but the sandwich treats them as zero. The effective number of independent observations in the factor dimension is T , not NT .

Step 8: SE ratio and coverage. Taking square roots,

$$\frac{\widehat{\text{se}}(\hat{\alpha}_{\text{TimeFE}})}{\text{se}(\hat{\alpha}_{\text{TimeFE}})} = \left(\frac{O(1/NT)}{O(1/N)}\right)^{1/2} = O\left(\frac{1}{\sqrt{T}}\right).$$

As $T \rightarrow \infty$ the reported confidence interval shrinks at rate $1/\sqrt{NT}$ around a correctly centered estimate whose true dispersion is $O(1/\sqrt{N})$ and invariant to T . Translating the ratio into coverage requires the estimation error to be asymptotically normal at its true scale, $(\hat{\alpha}_{\text{TimeFE}} - \alpha)/\text{se}(\hat{\alpha}_{\text{TimeFE}}) \Rightarrow \mathcal{N}(0, 1)$, which is Step 6 renormalized by $\sqrt{V/N}$. Writing $r_T \equiv \widehat{\text{se}}/\text{se} = O_p(T^{-1/2})$, the nominal 95% interval covers the truth with probability

$$\Pr(|\hat{\alpha}_{\text{TimeFE}} - \alpha| \leq 1.96 \widehat{\text{se}}) = \Pr(|Z| \leq 1.96 r_T) = 2\Phi(1.96 r_T) - 1 \rightarrow 2\Phi(0) - 1 = 0,$$

where Φ is the standard normal CDF; for small x , $2\Phi(x) - 1 = \sqrt{2/\pi} x + O(x^3)$, so coverage

vanishes at rate $O(T^{-1/2})$. The floor exists whenever $\sigma^2 > 0$; the restriction $\rho_{\lambda\mu} = 0$ serves only to switch off the bias channel of Lemma 1. $\square$ $\square$

B.4 What Variance Estimators Can and Cannot Price

[↔ Back to reference in main text](#)

This appendix provides the derivations underlying the comparison of variance estimators in Section 3.3, including their treatment of the variance component characterized in Lemma 3. The lemma is a statement about inference over the loading *superpopulation*: the draws (λ_i, μ_i) are part of the sampling experiment. The factor term $A_1 = \left(\sum_i \tilde{\mu}_i \tilde{\lambda}_i\right) S_F$ is one number for the realized panel, but it is a *sum of N unit-level terms*, and which variance estimators can price it is decided by whether they see those terms separately. Along time they cannot: every period's score $\sum_i \tilde{x}_{it} \hat{u}_{it}$ contains the same realized sum, so A_1 is a fixed displacement from the point of view of any estimator built from period-level scores. Across units they can: the unit-level score is $s_i = \tilde{\mu}_i \tilde{\lambda}_i S_F + r_i$ with $r_i = O_p(\sqrt{T})$, so the N scores are, to leading order, N independent draws of the exposure product, and their dispersion is the κ of the lemma. The taxonomy follows.

Heteroskedasticity-robust standard errors miss the floor for the reason proved in Appendix B.3: they treat the NT score contributions as independent and converge at the NT rate. *Clustering by time* and *Driscoll–Kraay* standard errors allow for within-period cross-sectional dependence, but do not capture the variance floor in this setting. This limitation does not follow from period-specific score orthogonality, which least squares does not impose: the normal equations deliver only $\sum_{i,t} \tilde{x}_{it} \hat{u}_{it} = 0$, not $\sum_i \tilde{x}_{it} \hat{u}_{it} = 0$ date by date. The reason is that each period-level score carries the same $\sum_i \tilde{\mu}_i \tilde{\lambda}_i$; its dispersion across dates reflects $\text{Var}(\tilde{F}_t^2)$, not κ , and is of smaller order. Both estimators spend degrees of freedom on that within-period covariation and deliver coverage below the robust benchmark (Table A.4).

Clustering by unit prices the floor. Write $\hat{u}_{it} = \tilde{\lambda}_i \tilde{F}_t + \tilde{\varepsilon}_{it} - (\hat{\alpha} - \alpha) \tilde{x}_{it}$, so that

$$s_i = \tilde{\mu}_i \tilde{\lambda}_i S_F + r_i, \quad r_i = \tilde{\lambda}_i \sum_t \tilde{\xi}_{it} \tilde{F}_t + \sum_t \tilde{x}_{it} \tilde{\varepsilon}_{it} - (\hat{\alpha} - \alpha) \sum_t \tilde{x}_{it}^2.$$

Under the weak-dependence conditions of Appendix B.3 the first two pieces of r_i are $O_p(\sqrt{T})$ and, since $\hat{\alpha} - \alpha = O_p(N^{-1/2})$ by Lemma 3, the third is $O_p(TN^{-1/2})$, against a leading term of order T . Hence $\sum_i s_i^2 = S_F^2 \sum_i (\tilde{\mu}_i \tilde{\lambda}_i)^2 [1 + O_p(T^{-1/2}) + O_p(N^{-1/2})]$. With finite fourth moments and $\rho_{\lambda\mu} = 0$, $N^{-1} \sum_i (\tilde{\mu}_i \tilde{\lambda}_i)^2 \rightarrow_p \kappa$; with $S_F/T \rightarrow_p \text{Var}(\tilde{F}_t)$ and $D/(NT) \rightarrow_p$

$\text{Var}(\tilde{x}_{it}),$

$$N \frac{\sum_i s_i^2}{D^2} \xrightarrow{p} \frac{\kappa [\text{Var}(\tilde{F}_t)]^2}{[\text{Var}(\tilde{x}_{it})]^2} = V,$$

which is the V of Lemma 3; nominal coverage under experiment (i) then follows from (8) and Slutsky's theorem, and the $\sqrt{T}$ ratio to the robust estimator from the corollary of Appendix B.3. Three remarks complete the proof. The terms in r_i are what an ordinary unit-cluster correction prices – within-unit serial dependence – and they are of smaller order here, so the estimator is right for the floor by construction rather than by accident. Because the normal equations impose $\sum_i s_i = 0$, a non-zero mean of the exposure product ($\rho_{\lambda\mu} \neq 0$) is absorbed into the bias rather than into the sandwich; the width stays $O(N^{-1/2})$ around a centre that is off by $O(1)$, so coverage is not restored (Table A.4, Panel C). And under the conditional experiment (ii) the same term is a displacement, so the unit-clustered interval is conservative there.

Two-way clustering combines the unit- and time-clustered variance estimators and subtracts the heteroskedasticity-robust estimator. Under the maintained assumptions, the latter two terms are asymptotically negligible relative to the unit-clustered component. Two-way and unit-clustered confidence intervals therefore have similar widths and coverage rates (Table A.4). Valid inference in the zero-bias case is thus possible without estimating the loadings, provided the variance estimator accounts for their cross-sectional sampling variation. The resulting interval width nevertheless remains of order $N^{-1/2}$ as T increases. Correctly modelling the factor structure removes this component of sampling uncertainty.

Remark 5 (The reference sampling experiment). Lemma 3 and the Monte Carlo coverage results refer to repeated samples in which loadings, factors and idiosyncratic shocks are redrawn. The variance floor reflects the cross-sectional sampling variation in the loading products. Conditional on the realized loadings, their sample covariance need not be zero even when the population covariance is zero; as T increases, this generates a persistent displacement of the Time FE estimate from α . The reported coverage rates therefore describe the unconditional sampling experiment and need not hold conditional on a particular realization of the loadings.

B.5 Distinct Aggregates: The Two-Factor Bias and Signal Deletion

[↔ Back to reference in main text](#)

The main text writes a single common factor F_t entering both equations, for clarity; nothing

in the framework requires the outcome and the regressor to load on the same aggregate (Section 3). This appendix relaxes that, letting

$$y_{it} = \alpha x_{it} + \lambda_i F_t^y + \gamma_i + \varepsilon_{it}, \quad x_{it} = \mu_i F_t^x + \xi_{it}, \quad (17)$$

with the outcome's error factor F^y and the regressor's factor F^x possibly distinct. Two things become visible that the one-factor form conceals: the omitted variable bias acquires a second determinant and a channel opens through which IFE and time dummies differ even when that bias is zero.

The bias in two factors. Repeating the derivation of Lemma 1 under (17), the time fixed effects estimator satisfies

$$\text{plim}(\hat{\alpha}_{\text{TimeFE}} - \alpha) = \frac{\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) \text{Cov}(\tilde{F}_t^x, \tilde{F}_t^y)}{\text{Var}(\tilde{x}_{it})}, \quad (18)$$

where $\text{Cov}(\tilde{F}_t^x, \tilde{F}_t^y)$ is the temporal covariance of the two demeaned aggregates. The one-factor formula (3) is the special case $F^x = F^y$, for which this reduces to $\text{Var}(\tilde{F})$. The bias is now the product of *two* alignments – the cross-sectional covariance of the loadings and the temporal covariance of the aggregates – and it vanishes if *either* is zero. In particular, the bias is zero whenever the regressor's aggregate is orthogonal to the outcome's error factor, however strongly the loadings covary. This is why the plug-in omitted variable term is small in the application: the estimated inflation and slack factors are only weakly aligned (correlation 0.305), so $\text{Cov}(\tilde{F}^x, \tilde{F}^y)$ is small and holds the textbook channel near zero irrespective of $\text{Cov}(\mu, \lambda)$.

The corollary below collects the formal two-aggregate statements that accompany Lemma 2 in the main text: the two-factor form of the bias, in which the conflation of the lemma and the omitted variable term appear side by side, and the defactoring result that underlies Step 1 of the two-step estimator of Remark 2.

Corollary 1 (Two aggregates: bias and defactoring). *Let the outcome and regressor follow the two-factor environment (17) with the component responses of (4), under the conditions of Lemma 2, and let w_e be the synchronized-component weight of (5). Then:*

(i) (Bias.) *For the time fixed effects estimator,*

$$\text{plim } \hat{\alpha}_{\text{TimeFE}} = w_{\text{TFE}} \alpha_c + (1 - w_{\text{TFE}}) \alpha_x + \frac{\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) \text{Cov}(\tilde{F}_t^x, \tilde{F}_t^y)}{\text{Var}(\tilde{x}_{it})}, \quad (19)$$

so the omitted variable term of Lemma 1 survives as the product of two alignments – the cross-sectional covariance of the loadings and the temporal covariance of the aggregates – and vanishes only if either is zero. At $\alpha_c = \alpha_x = \alpha$ the first two terms collapse to α and the display reduces to the bias formula (18); with component-specific responses the same expression survives as an additive term in the probability limit rather than as a deviation from a single α .

(ii) (Defactoring.) If the regressor is first stripped of its common component – x_{it} replaced by $\xi_{it} = x_{it} - \mu_i F_t^x$ – then $w_e = 0$ by the maintained independence of ξ_{it} from the loadings and the factor paths, and $\text{plim } \hat{\alpha} = \alpha_x$ under idiosyncratic exogeneity together with that independence, for any absorption scheme: the outcome’s factor term is orthogonal to ξ_{it} by the same independence.

Proof. The Frisch–Waugh–Lovell argument in Appendix B.2, generic in the absorption scheme e , delivers the conflation identity (6). Part (ii) is the case $\tilde{x}_{it}^{(e)} = \xi_{it}$, for which the covariance in (5) is zero. Part (i) adds to the identity the contribution of the unabsorbed outcome-factor term $\lambda_i F_t^y$, evaluated for the two-way transformation in equation (18); the two contributions are additive by the linearity of the projection. $\square$

Identification and the retained signal. Consistency of IFE is governed by the rank condition (Assumption 4), $\inf_F D(F) > 0$. As noted there, this is *robustly* secured by idiosyncratic regressor variation $\text{Var}(\xi_{it}) > 0$: no factor process can absorb ξ , so the infimum is bounded below by its variance. The common component $\mu_i F_t^x$ does not secure the condition in this robust sense – the candidate $F = F^x$ drives D to zero – and identifies α only under the weaker requirement that D be positive at the true factor (Bai 2009), with the rank correctly specified.

The economic content, though, runs mostly through that common component. Under correct rank IFE removes only the outcome factor $\lambda_i F_t^y$, and so *retains* $\mu_i F_t^x$ as identifying variation to the extent that F^x is not aligned with by F^y . A full set of time dummies, by contrast, removes the cross-sectional average at each date and thereby deletes the *synchronized* movement $\bar{\mu} F_t^x$ – the economy-wide slack episodes – retaining only the loading dispersion and the idiosyncrasy. When the two aggregates coincide the distinction collapses: IFE removes $\mu_i F_t^x$ along with the error factor and both estimators identify from ξ . When they are only weakly aligned, as here, the synchronized slack movements survive under IFE and are discarded under time dummies. This is the precise sense in which IFE “retains the signal” that time dummies delete, and it is why the two estimators can diverge even where (18) is zero.

Remark 6 (Signal deletion when the aggregates are distinct). The channel described here is stated formally in the main text as Lemma 2 (Failure II); this remark is the appendix discussion of its mechanism. The retained signal shifts the *estimand*, not merely the precision, once the outcome responds differently to the synchronized and idiosyncratic components of the regressor. Take first the zero-bias benchmark in which the omitted variable term of (18) vanishes – that is, $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i) = 0$ or $\text{Cov}(\tilde{F}_t^x, \tilde{F}_t^y) = 0$. There, when α is constant, IFE and time dummies both identify it, and deleting the synchronized component of slack only inflates the variance of the time fixed effects estimate. Outside that benchmark the contrast drawn here does not arise in isolation: with both covariances non-zero, time dummies remain biased even under component-invariant responses, and the estimand channel below adds to that bias rather than standing in for it. When the responses differ across components – write α_c for the response to the synchronized movement $\mu_i F_t^x$ and α_x for the response to the idiosyncratic movement ξ_{it} – each estimator recovers a variance-weighted average of the two over the variation it uses. Time dummies delete the average of the synchronized movement, so their estimand collapses toward α_x ; IFE retains that movement whenever the regressor’s aggregate is not the outcome’s, so its estimand keeps a substantial weight on α_c . The channel requires no loading covariance and survives $\text{Cov}(\mu, \lambda) = 0$; it is switched on by the combination of a dominant, independently-driven regressor factor and component-specific responses, both directly measurable – the former from the within-variance share of the first factor of x and its correlation with the outcome factor, the latter from the component slopes estimated in Section 7.4. The accounting exercise of Section 7.4 finds this channel, operating through component-specific slopes rather than threshold reweighting, to be the dominant priced force in the time-FE–IFE gap in the euro area panel, well ahead of the omitted variable term, which is negligible and unstable in sign; the remainder is the co-movement of inflation with the quarter averages of tightness that the period dummies delete.

B.6 Proof of Proposition 1

[*↔ Back to reference in main text*](#)

The proof follows closely the structure of Bai (2009). Steps 1 and 2 verify that the environment of Section 3 is a special case of his model and that Assumptions 2–5 deliver his conditions, at which point his consistency and asymptotic-normality results apply directly and are not reproved here. The paper’s own contribution is Step 3, which asks what general-equilibrium feedback does to those conditions and shows that Assumption 1 is what keeps them intact.

Proposition 1 (IFE Identification). *Under Assumption 1 and Assumptions 2–5, with $N, T \rightarrow \infty$ such that $T/N \rightarrow \varrho \in (0, \infty)$, the IFE estimator $\hat{\alpha}_{IFE}$ is:*

- (i) *consistent: $\hat{\alpha}_{IFE} \xrightarrow{P} \alpha$;*
- (ii) *$\sqrt{NT}$ -asymptotically normal, with variance as characterised in Bai (2009, Theorem 3) (with the bias correction of his Theorem 4 under heteroskedasticity or cross-sectional correlation when $T/N \rightarrow \varrho > 0$);*
- (iii) *strictly preferred to $\hat{\alpha}_{TimeFE}$, in the sense that $|\text{plim}(\hat{\alpha}_{TimeFE} - \alpha)| > 0$ whenever $\rho_{\lambda\mu} \neq 0$, while the nominal 95% coverage of $\hat{\alpha}_{TimeFE}$ falls below 0.95 whenever $\sigma^2 > 0$.*

Proof. We proceed in four steps corresponding to the three claims plus the GE extension.

Step 1: Mapping to Bai (2009). Write the DGP as

$$y_{it} = \alpha x_{it} + \lambda_i F_t + \underbrace{\gamma_i + \varepsilon_{it}}_{=: e_{it}}, \quad x_{it} = \mu_i F_t + \xi_{it}.$$

This is a special case of Bai’s equation (1), $y_{it} = x'_{it}\beta + \lambda'_i F_t + e_{it}$, with scalar $\beta = \alpha$, scalar factor F_t , loading λ_i , and idiosyncratic error e_{it} . Unit fixed effects γ_i are handled by the within transformation, which partials them out before the factor structure is estimated; this is the route we take and it is the one Bai’s framework accommodates directly, since he treats additive and interactive effects jointly. The alternative of folding γ_i into the loading deserves a word of caution, because it does not work as stated for a scalar factor: an arbitrary constant γ_i cannot be represented as $\lambda_i F_t$ when F_t is non-constant. Absorption requires augmenting the factor vector with a constant component, so that the model carries a factor $(1, F_t)'$ with loading (γ_i, λ_i) , which raises the nominal rank by one. Nothing downstream depends on the choice, and we use the within route throughout. The critical feature is that x_{it} shares the factor F_t with the composite error $\lambda_i F_t + e_{it}$, which Bai explicitly accommodates through his rank condition.

Step 2: Verifying Bai’s conditions via Assumptions 2–5.

- (a) *Rank condition (Bai Assumption A).* Bai requires $\mathbb{E}\|X_{it}\|^4 \leq M$ and $\inf_F D(F) > 0$, where $D(F)$ is formed from the regressors after projecting out the candidate factors and cross-sectional averages. By Assumption 4, $\text{Var}(\xi_{it}) > 0$; since $x_{it} = \mu_i F_t + \xi_{it}$, the projection leaves ξ_{it} -variation in every unit, which is Bai’s own sufficient condition for his Assumption A under exactly this design.

- (b) *Factor and loading strength (Bai Assumption B)*. Assumption 3(i) gives $T^{-1} \sum_t F_t^2 \xrightarrow{p} \Sigma_F > 0$, Bai's requirement B(i) that $T^{-1} F' F \rightarrow \Sigma_F$ positive definite; Assumption 3(ii) gives $N^{-1} \Lambda' \Lambda \rightarrow \Sigma_\lambda > 0$, his B(ii).
- (c) *Weak error dependence (Bai Assumption C)*. Assumption 5 imposes bounded cross-sectional covariance eigenvalues, absolutely summable autocovariances and bounded eighth moments on $e_{it} = \gamma_i + \varepsilon_{it}$ after the unit effect γ_i is absorbed by the within transformation; this is Bai's Assumption C.
- (d) *Exogeneity (Bai Assumption D)*. Bai requires e_{it} independent of X_{js} , λ_j and F_s for all j, s . Idiosyncratic exogeneity ($\mathbb{E}[\xi_{is}\varepsilon_{it}] = 0$ for all s, t) together with the independence of ε_{it} from the loadings and $\{F_t\}$ (Section 3) delivers this; it rules out lagged dependent variables in x_{it} , which the specification does not contain.

Under (a)–(d), Bai (2009, Proposition 1) delivers Part (i), consistency. Under the additional CLT condition of his Assumption E – satisfied here because the weak-dependence structure of Assumption 5 makes the score's long-run variance finite and positive definite – his Theorem 3 delivers Part (ii), $\sqrt{NT}$ -asymptotic normality.

Step 3: GE feedback does not invalidate the rank condition (role of Assumption 1). In the presence of general-equilibrium feedback operating through the *weighted* aggregate of Assumption 1, the outcome equation becomes

$$y_{it} = \alpha x_{it} + \lambda_i F_t + \psi \theta_i \bar{X}_t^w + \gamma_i + \varepsilon_{it}, \quad \bar{X}_t^w = \sum_j s_j x_{jt} = \bar{\mu}^w F_t + \bar{\xi}_t^w,$$

where $\bar{\mu}^w \equiv \sum_j s_j \mu_j$ and $\bar{\xi}_t^w \equiv \sum_j s_j \xi_{jt}$, with $\sum_j s_j = 1$. Substituting,

$$y_{it} = \alpha x_{it} + \underbrace{(\lambda_i + \psi \theta_i \bar{\mu}^w)}_{\tilde{\lambda}_i} F_t + \psi \theta_i \bar{\xi}_t^w + \gamma_i + \varepsilon_{it}.$$

The term $\psi \theta_i \bar{\mu}^w F_t$ is absorbed into a rescaled factor loading $\tilde{\lambda}_i$; Bai's conditions (a)–(d) continue to hold with λ_i replaced by $\tilde{\lambda}_i$ (since $\sigma_{\tilde{\lambda}}^2 > 0$ follows from $\sigma_\lambda^2 > 0$ and $\sigma_\theta^2 > 0$, and $\bar{\mu}^w = O(1)$ because the weights sum to one). The residual GE term $\psi \theta_i \bar{\xi}_t^w$ has two potential effects, and granularity controls both.

(3a) *Endogeneity bias*. By independence of ξ_{it} across units,

$$\text{Cov}(\xi_{it}, \bar{\xi}_t^w) = s_i \text{Var}(\xi_{it}) = s_i \sigma_\xi^2,$$

which is the unit-level bound (10): unit i 's idiosyncratic variation contaminates the aggregate in proportion to its own share, and the contamination is *heterogeneous* across units, of order $O(s_i)$. Aggregating over the panel, the pooled IFE moment condition inherits the cross-sectional average of these terms over the post-projection variance of the regressor, which by Step 2(a) is σ_ξ^2 :

$$\text{Bias}_{\text{GE}} = \frac{\text{Cov}(\xi_{it}, \psi\theta_i\bar{\xi}_t^w)}{\sigma_\xi^2} = \psi N^{-1} \sum_i \mathbb{E}[\theta_i] s_i, \quad \left| N^{-1} \sum_i \mathbb{E}[\theta_i] s_i \right| \leq \bar{\theta} \max_i s_i,$$

with $\bar{\theta} \equiv N^{-1} \sum_i |\mathbb{E}[\theta_i]|$. Hence $|\text{Bias}_{\text{GE}}| \leq |\psi| \bar{\theta} \max_i s_i = O(\max_i s_i)$, which vanishes under Assumption 1 and is the operational bound (11). Under equal weights $s_i = N^{-1}$ this reduces to the parametric rate $\text{Bias}_{\text{GE}} = \psi\bar{\theta}/N = O(1/N)$, with $\bar{\theta} \equiv N^{-1} \sum_i \mathbb{E}[\theta_i]$ the signed mean loading.

(3b) *Factor strength of the GE residual.* The deeper threat is not the average bias but the possibility that $\psi\theta_i\bar{\xi}_t^w$ constitutes an *additional common factor*: it is a single time series $\bar{\xi}_t^w$ loading on every unit through $\psi\theta_i$, exactly the structure that Bai's Assumption C rules out for the error term and that would raise the true rank above the estimated one. Its strength is governed by

$$\text{Var}(\bar{\xi}_t^w) = \sum_j s_j^2 \sigma_\xi^2 = \text{HHI} \cdot \sigma_\xi^2,$$

so the GE residual is a common component whose variance is proportional to the Herfindahl index. Vanishing variance is not by itself the right criterion, however, and it is worth being careful here. Bai's Assumption C bounds the eigenvalues of the *cross-sectional covariance matrix* of the composite error, and the GE term contributes the rank-one matrix $\psi^2 \theta\theta' \text{Var}(\bar{\xi}_t^w)$, whose single non-zero eigenvalue is

$$\psi^2 \|\theta\|^2 \text{Var}(\bar{\xi}_t^w) = \psi^2 \left(\sum_i \theta_i^2 \right) \text{HHI} \cdot \sigma_\xi^2 = O(N \cdot \text{HHI})$$

when the loadings θ_i are non-degenerate, since then $\sum_i \theta_i^2 = O(N)$. The eigenvalue therefore scales with $N \cdot \text{HHI}$, not with HHI: driving the variance to zero at rate HHI while the cross-section grows at rate N leaves the eigenvalue bounded only if the two rates offset. The condition we need is accordingly the sharper

$$N \cdot \text{HHI} = O(1), \tag{20}$$

which implies $\text{HHI} \rightarrow 0$ but is not implied by it. Under (20) the GE residual is a bounded-eigenvalue disturbance rather than a genuine factor, Bai's Assumption C holds for the com-

positive error $\psi\theta_i\bar{\xi}_t^w + \varepsilon_{it}$, and the rank condition is preserved, so Part (i) continues to hold, and Part (ii) holds with the feedback term of (3a) added to the centre of the limiting distribution, a shift of order $|\psi|\sqrt{T/N}$.

The strengthening costs nothing in the application, and it is directly checkable. Under equal weights $\text{HHI} = 1/N$ identically, so $N \cdot \text{HHI} = 1$. Under official EU Commission output weights the harmonized sample has $\text{HHI} = 0.022$ and $N \cdot \text{HHI} \approx 2.7$. Both are bounded, and comfortably so; the binding case is the weighted one. Conversely, when concentration keeps HHI bounded away from zero, the residual feedback component can constitute an additional pervasive factor. Rank-one IFE is then generally misspecified. Allowing a second factor can accommodate this component.

Step 4: Part (iii) – strict preference over Time FE. *Bias channel.* By Lemma 1,

$$\text{plim}(\hat{\alpha}_{\text{TimeFE}} - \alpha) = \frac{\rho_{\lambda\mu}\sigma^2(1 - N^{-1}) \text{Var}(\tilde{F}_t)}{\text{Var}(\tilde{x}_{it})}.$$

This is strictly non-zero whenever $\rho_{\lambda\mu} \neq 0$ and $\sigma^2 > 0$. By Part (i), $\text{plim}(\hat{\alpha}_{\text{IFE}} - \alpha) = 0$ under the same conditions, so IFE strictly dominates in probability limit.

Inference channel. By Lemma 3, even when $\rho_{\lambda\mu} = 0$, the nominal 95% coverage of $\hat{\alpha}_{\text{TimeFE}}$ satisfies

$$\text{Coverage}_T \rightarrow 0 \quad \text{as } T \rightarrow \infty,$$

whenever $\sigma^2 > 0$, because the robust standard error understates the true dispersion by a factor $O(1/\sqrt{T})$. By Part (ii), $\hat{\alpha}_{\text{IFE}}$ is $\sqrt{NT}$ -normal with a consistently estimated variance, so its coverage converges to 0.95 when that shift is negligible. The nominal coverage of $\hat{\alpha}_{\text{TimeFE}}$ therefore falls strictly below 0.95 for all large T whenever $\sigma^2 > 0$, which by Assumption 2 holds. □ □

Remark 7 (Technical conditions in practice). Four implementation details connect Proposition 1 to estimation. (i) *Normalization.* The pair (Λ, F) is identified only up to an invertible $r \times r$ rotation; estimation imposes the principal components convention $T^{-1}F'F = I_r$ with $\Lambda'\Lambda$ diagonal. The slope α – the object of interest – is invariant to the choice of normalization, so none of the results depend on it. (ii) *Fixed effects.* The additive effects $\gamma_i + \tau_j$ are absorbed by the within transformation inside each iteration of the least-squares algorithm; Bai (2009) treats additive and interactive effects jointly and shows the combined estimator retains the properties of Proposition 1. (iii) *Residualized-regressor rank.* The rank condition of Assumption 4 is a statement about the regressor *after* the estimated factor space is pro-

jected out: it fails not when x_{it} is strongly factor-driven per se, but when nothing is left once the factor is removed ($\text{Var}(\xi_{it}) = 0$), in which case no within-unit variation identifies α under any estimator. *(iv) Generated factors and standard errors.* The variance in Proposition 1(ii) already accounts for the estimation of F and Λ : factor estimation does not slow the $\sqrt{NT}$ rate. In the joint limit $T/N \rightarrow \varrho > 0$, heteroskedasticity or cross-sectional correlation in the errors shifts the limiting distribution by $O(1)$ bias terms; the application reports clustered and bootstrap standard errors as safeguards.

C Complements to the Method

[↔ Back to reference in main text](#)

This appendix reports the additional simulations summarized in Section 5, using the calibration described there.

Experiment G7: general-equilibrium feedback. Table A.1 introduces an additional aggregate, $M_t = \rho_M M_{t-1} + \psi \bar{x}_t + u_t^M$, which enters the outcome through loadings $\theta_i \sim \mathcal{N}(1, 1)$. Feedback from $\bar{x}_t$ makes this component correlated with the regressor. Rank-one IFE leaves an omitted common component: coverage falls from 0.96 in the benchmark to 0.25–0.30 in the feedback specifications, although the mean estimates remain close to unity. Rank-two IFE yields approximately unbiased estimates and coverage of about 0.90 even at $\psi = 0.5$.

The experiment illustrates the consequences of underestimating the factor rank. An omitted pervasive factor can distort inference even when the coefficient bias is small. Rank-selection criteria and residual factor diagnostics can help detect such misspecification; allowing an additional factor accommodates the feedback component in this design.

Experiment G8: granularity and feedback. We consider feedback through the equal-weighted aggregate of the regressor:

$$\begin{aligned} y_{it} &= \alpha x_{it} + \lambda_i F_t + \psi \theta_i \bar{x}_t + \gamma_i + \varepsilon_{it}, \\ x_{it} &= \mu_i F_t + \xi_{it}, \quad \bar{x}_t = \frac{1}{N} \sum_i x_{it}, \end{aligned} \tag{21}$$

with $\theta_i \sim \mathcal{N}(1, 1)$ and $\rho_{\lambda\mu} = 0$. The feedback term decomposes as

$$\psi \theta_i \bar{x}_t = \underbrace{\psi \theta_i \bar{\mu} F_t}_{\text{collinear with } F_t} + \underbrace{\psi \theta_i \bar{\xi}_t}_{\text{residual feedback}}. \tag{22}$$

TABLE A.1: General Equilibrium Feedback

Estimator	<i>GE feedback ψ</i>		
	$\psi = 0$	$\psi = 0.3$	$\psi = 0.5$
Time FE	0.999 {0.30}	1.002 {0.35}	1.002 {0.36}
IFE (rank 1)	1.000 {0.96}	0.992 {0.25}	0.995 {0.30}
IFE (rank 2)	1.000 {0.96}	1.000 {0.94}	1.000 {0.91}
<i>Condition: $\rho_{\lambda\mu} = 0$. {Coverage}</i>			

Notes: 500 replications. $\sigma = 1$, $\rho_{\lambda\mu} = 0$, $N = 250$, $T = 100$. A second factor $M_t = \rho_M M_{t-1} + \psi \bar{x}_t + u_t^M$ enters inflation with unit-specific loadings $\theta_{ij} \sim \mathcal{N}(1, 1)$: aggregate inflation feeds back into the regressor, so once $\psi > 0$ the aggregate $\bar{x}_t$ reflects both F_t and M_t . Rank-one IFE is then fitting one factor to a two-factor world and leaves $\theta_{ij} M_t$ in the residual, which is why its coverage deteriorates sharply while its point estimate moves comparatively little. IFE rank 2 jointly estimates both factors and remains unbiased with correct coverage. Coverage in braces.

The first component is absorbed into the loading on F_t . Under cross-sectional independence, the second has variance proportional to $1/N$, and its covariance with ξ_{it} is $\psi \theta_i \sigma_\xi^2 / N$. Increasing N therefore reduces the feedback-induced violation of the idiosyncratic moment condition.

Table A.2 varies N from 10 to 400 at three feedback strengths. Rank-one IFE exhibits upward bias and undercoverage in the smallest panels. Both diminish as N increases; by $N = 200$, the estimates are close to unity and coverage is near nominal across the feedback strengths considered. The decline in bias is consistent with a $1/N$ rate. Time FE estimates remain close to unity, but coverage based on heteroskedasticity-robust standard errors remains low throughout, reflecting the separate inference problem associated with unmodelled factor loadings.⁴²

Experiment G5, continued: inference for the synchronized slope. Table A.3 extends the second-step IV experiment. The common tightness factor is $F_t^x = z_t + h_t$, where z_t is excluded from the outcome and h_t enters with loading $\theta\mu_i$. The simulations reproduce

⁴²With $\rho_{\lambda\mu} = \pm 0.3$, Time FE also exhibits coefficient bias that persists as N increases. IFE remains approximately unbiased, with coverage increasing from about 0.65 at $N = 10$ to 0.95 at $N = 200$; results are available on request.

TABLE A.2: Granularity Dividend: GE Bias of IFE Vanishes as N Grows

N	HHI	<i>GE feedback strength ψ</i>		
		$\psi = 0$	$\psi = 0.3$	$\psi = 0.5$
<i>Panel A: Time FE</i>				
10	0.100	0.999	0.995	0.980
		{0.28}	{0.24}	{0.27}
25	0.040	0.985	1.003	1.002
		{0.30}	{0.29}	{0.27}
50	0.020	1.002	0.996	1.004
		{0.27}	{0.27}	{0.28}
100	0.010	1.000	1.000	1.003
		{0.32}	{0.28}	{0.25}
200	0.005	0.998	0.998	0.999
		{0.31}	{0.27}	{0.34}
400	0.0025	0.999	1.000	1.003
		{0.28}	{0.31}	{0.27}
<i>Panel B: IFE</i>				
10	0.100	1.001	1.028	1.048
		{0.93}	{0.82}	{0.65}
25	0.040	1.000	1.012	1.019
		{0.95}	{0.92}	{0.85}
50	0.020	1.001	1.006	1.009
		{0.94}	{0.94}	{0.93}
100	0.010	1.000	1.003	1.005
		{0.96}	{0.94}	{0.93}
200	0.005	1.000	1.001	1.003
		{0.94}	{0.95}	{0.94}
400	0.0025	1.000	1.001	1.001
		{0.93}	{0.94}	{0.95}

Notes: Each cell reports the mean point estimate (top) and the coverage of nominal 95% confidence intervals (in braces) over 500 replications; the true slope is $\alpha = 1$. Loadings are independent ($\rho_{\lambda\mu} = 0$), shutting the omitted-variable channel so that any bias is attributable to feedback alone. The $\psi = 0$ column is a validity benchmark with no feedback.

the projection coefficient

$$\text{plim } \hat{\alpha}_c^{\text{OLS}} = \alpha_c + \theta \frac{\text{Var}(h)}{\text{Var}(F^x)}.$$

Instrumenting with $\mu_i z_t$ removes this confounding: median IV estimates lie within 0.004 of α_c in every specification. Replacing the common component and loadings with principal component estimates introduces a small downward bias, with $\text{corr}(\hat{c}, c) = 0.99$.

Inference depends on the T aggregate observations. With a persistent level instrument, the date-level score is serially correlated, and date-clustered standard errors yield coverage of 0.63–0.67 at $T = 100$ in the specifications with confounding. Increasing T to 1,000 does not resolve this distortion. A Bartlett kernel with fixed- b critical values or an equal-weighted cosine estimator improves coverage to 0.86–0.89 at $T = 100$ and approximately 0.94 at $T \geq 500$. Undercoverage nevertheless remains when factor persistence is raised to the empirical value of 0.96.

Using the innovation to z_t instead of its level eliminates score serial correlation in this design. Time-clustered coverage is then 0.91–0.96 across the persistence parameters and panel lengths considered. Instrument relevance becomes more restrictive: at persistence 0.95, the first-stage F -statistic is approximately 6 at $T = 100$ and 14 at $T = 500$ (replication files `v14_A` and `v14_B`).

TABLE A.3: Two-step estimator’s second step, IV step for the common-component slope

Design			OLS on c_{it}		IV			IV, feasible	
α_c	θ	Derived target	Mean	$-\alpha_c$	Median	Mean	Cov. _{cl}	Median	Mean
1.00	0.00	1.000	1.000	0.0002	1.000	1.000	0.95	1.000	1.000
1.00	0.50	1.248	1.248	0.2482	0.997	0.969	0.67	0.986	0.957
1.00	1.00	1.497	1.497	0.4972	1.000	0.950	0.67	0.988	0.928
2.50	0.00	2.500	2.500	0.0001	2.501	2.500	0.93	2.470	2.469
2.50	0.50	2.752	2.752	0.2525	2.499	2.473	0.63	2.457	2.430
2.50	1.00	3.002	3.002	0.5021	2.503	2.450	0.67	2.448	2.396

Notes: The aggregate slack impulse splits as $F_t^x = z_t + h_t$, independent AR(1) processes with $\text{Var}(h)/\text{Var}(F^x) = 0.5$. z is excluded from the outcome; h enters it with loading $\theta\mu_i$, proportional to the regressor loadings. $\theta = 0$ is the no-confound benchmark. Derived target is the resulting OLS plim, $\alpha_c + 0.5\theta$. Instrument $\mu_i z_t, \hat{\mu}_i z_t$ (feasible); first-stage F between 129 and 137 throughout. The IV is a ratio of time-series covariances, so its mean sits about 0.05θ below α_c at $T = 100$ even when the estimator is exact; both are reported so the drift is visible. Cov._{cl} uses standard errors clustered by date. $R = 500$.

Experiment G9: variance estimators and coverage. Table A.4 compares variance estimators as T increases, holding the coefficient estimates fixed. At $\rho_{\lambda\mu} = 0$, Time FE is

approximately unbiased, but heteroskedasticity-robust, time-clustered and Driscoll–Kraay standard errors all yield declining coverage. Between $T = 20$ and $T = 200$, coverage falls from 0.68 to 0.22 under robust inference, from 0.51 to 0.14 under time clustering, and from 0.47 to 0.18 under Driscoll–Kraay.

Unit clustering restores nominal coverage by accounting for the variance floor. Its standard error remains approximately 0.03 across panel lengths, matching the Monte Carlo standard deviation. The ratio to the robust standard error increases from 2.2 to 6.5, consistent with the $\sqrt{T}$ rate. Two-way clustering produces similar intervals and coverage of 0.94–0.96, as the unit-clustered component dominates (Appendix B.4). IFE maintains coverage of 0.94–0.95 while its standard error declines with T .

At $\rho_{\lambda\mu} = 0.5$ (Panel C), all variance estimators yield zero coverage because the Time FE coefficient is biased by approximately 0.25. Unit clustering therefore corrects inference in the zero-bias design but does not address coefficient bias. The appropriate covariance estimator differs in Experiment G5, where the synchronized IV coefficient is identified from aggregate time-series variation and inference must account for dependence in the date-level score.

Experiment G10: loading heterogeneity. Table A.5 varies $\sigma \in \{0.5, 1.0, 2.0\}$ with $\rho_{\lambda\mu} = 0$. Time FE remains approximately unbiased, but coverage falls from 0.52 at $\sigma = 0.5$ to 0.30 at $\sigma = 2.0$. IFE estimates remain within 0.001 of unity, with coverage close to the nominal level throughout.

The next experiment operationalizes the group×time dummy ladder; the corollary it tests is stated first.

Corollary 2 (Group×Time Fixed Effects). *Maintain the DGP of equations (1)–(2) and Assumptions 2–5. Let $g : \{1, \dots, N\} \rightarrow \{1, \dots, G\}$, $G \geq 1$, be an arbitrary partition of units into equal-sized cells of size $n_g = N/G$. Decompose the loadings hierarchically:*

$$\lambda_i = \bar{\lambda} + \lambda_{g(i)}^g + \lambda_i^w, \quad \mu_i = \bar{\mu} + \mu_{g(i)}^g + \mu_i^w,$$

where the g -layer $\{(\lambda_g^g, \mu_g^g)\}_{g=1}^G$ and within-group layer $\{(\lambda_i^w, \mu_i^w)\}_{i=1}^N$ are mean-zero, mutually independent across layers and units, with within-group cross-loading correlation $\rho_w = \text{Corr}(\mu_i^w, \lambda_i^w)$ and dispersions $\sigma_{\mu,w}, \sigma_{\lambda,w} > 0$. Let $\hat{\alpha}_{g \times t}$ be the within-estimator that absorbs unit fixed effects and $g \times t$ fixed effects jointly.

TABLE A.4: Which Variance Estimator Prices the Floor: Change in T

Estimator / SE	<i>Time periods T</i>			
	$T = 20$	$T = 50$	$T = 100$	$T = 200$
<i>Panel A: $\rho_{\lambda\mu} = 0$. Mean {coverage}</i>				
Time FE, HC robust	0.998 {0.68}	1.002 {0.44}	1.000 {0.34}	1.001 {0.22}
Time FE, clustered by date	0.998 {0.51}	1.002 {0.30}	1.000 {0.19}	1.001 {0.14}
Time FE, Driscoll–Kraay	0.998 {0.47}	1.002 {0.35}	1.000 {0.27}	1.001 {0.18}
Time FE, clustered by unit	0.998 {0.96}	1.002 {0.94}	1.000 {0.95}	1.001 {0.96}
Time FE, two-way clustered	0.998 {0.96}	1.002 {0.94}	1.000 {0.95}	1.001 {0.95}
IFE (rank 1)	1.000 {0.94}	1.000 {0.95}	1.000 {0.94}	1.000 {0.95}
<i>Panel B: $\rho_{\lambda\mu} = 0$. Median standard error</i>				
Time FE, HC robust	0.0147	0.0093	0.0067	0.0048
Time FE, clustered by date	0.0102	0.0063	0.0044	0.0032
Time FE, Driscoll–Kraay	0.0101	0.0075	0.0057	0.0044
Time FE, clustered by unit	0.0320	0.0317	0.0319	0.0311
Time FE, two-way clustered	0.0311	0.0313	0.0316	0.0309
IFE (rank 1)	0.0115	0.0061	0.0041	0.0028
s.d. of $\hat{\alpha}_{\text{TimeFE}}$, Monte Carlo	0.0335	0.0341	0.0308	0.0317
Ratio, unit / HC	2.18	3.40	4.78	6.52
<i>Panel C: $\rho_{\lambda\mu} = 0.5$ (biased centre). Mean {coverage}</i>				
Time FE, HC robust	1.239 {0.00}	1.247 {0.00}	1.246 {0.00}	1.246 {0.00}
Time FE, clustered by date	1.239 {0.00}	1.247 {0.00}	1.246 {0.00}	1.246 {0.00}
Time FE, Driscoll–Kraay	1.239 {0.00}	1.247 {0.00}	1.246 {0.00}	1.246 {0.00}
Time FE, clustered by unit	1.239 {0.00}	1.247 {0.00}	1.246 {0.00}	1.246 {0.00}
Time FE, two-way clustered	1.239 {0.00}	1.247 {0.00}	1.246 {0.00}	1.246 {0.00}

Notes: DGP: $\alpha = 1$, $\rho_F = 0.80$, $\rho_\xi = 0.80$, $N = 250$, $\sigma = 1$. The Time FE rows report the same point estimate under different variance estimators; only the standard errors and coverage differ. Driscoll–Kraay lag $L = \lfloor 4(T/100)^{2/9} \rfloor$, computed on Frisch–Waugh residuals. Time FE rows: 500 replications; IFE row: 500 replications. Panel C has no $T = 20$ cell.

TABLE A.5: Generic Panel: Performance vs. Loading Heterogeneity σ ($N = 250$, $T = 100$)

Estimator	Loading heterogeneity σ		
	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 2.0$
Time FE	0.999 {0.52}	1.002 {0.30}	1.001 {0.30}
IFE	1.000 {0.94}	1.000 {0.96}	1.000 {0.95}
<i>Condition: $\rho_{\lambda\mu} = 0$. {coverage}</i>			

Notes: $\alpha = 1.0$, $\rho_F = 0.80$, $\rho_\xi = 0.80$, $N = 250$, $T = 100$. $R = 500$.

(i) **Bias.**

$$\text{plim}_{N,T \rightarrow \infty}(\hat{\alpha}_{g \times t} - \alpha) = \frac{\rho_w \sigma_{\mu,w} \sigma_{\lambda,w} \left(1 - \frac{G}{N}\right) \text{Var}(\tilde{F}_t)}{\text{Var}(\check{x}_{it})}, \quad (23)$$

where $\tilde{F}_t = F_t - \bar{F}$ and $\check{x}_{it}$ denotes x_{it} after the unit-plus- $g \times t$ within-transformation. The g -layer is annihilated exactly; only the within-group covariance $\rho_w \sigma_{\mu,w} \sigma_{\lambda,w}$ survives. The finite-cell correction $(1 - G/N) = (1 - 1/n_g)$ is the exact analog of the $(1 - 1/N)$ term in Lemma 1 with N replaced by the cell size n_g . At $G = 1$ (23) reduces to Lemma 1 exactly.

(ii) **Inference.** Whenever $\sigma_w := \sqrt{\sigma_{\mu,w} \sigma_{\lambda,w}} > 0$,

$$\frac{\widehat{\text{se}}(\hat{\alpha}_{g \times t})}{\text{se}(\hat{\alpha}_{g \times t})} = O\left(\frac{1}{\sqrt{T}}\right), \quad (24)$$

so the nominal 95% coverage of $\hat{\alpha}_{g \times t}$ decreases to zero as $T \rightarrow \infty$. The floor level scales with $\sigma_{\mu,w}^2 \sigma_{\lambda,w}^2$ – equivalently σ_w^4 under the definition above – since the object entering the variance is the product $\tilde{\mu}_i \tilde{\lambda}_i$, whose second moment carries both dispersions squared. It is therefore smaller at finer partitions; the rate $O(1/\sqrt{T})$ in (24) is identical at every rung of the dummy ladder. Interacting dummies shrinks the bias but does not slow the coverage collapse.

(iii) **Exhaustion.** At $G = N$ each cell contains a single unit, so unit $\times$ time effects span all NT observations and the regression has no residual; the ladder terminates. For every feasible $G < N$ at least two units share a cell, so $\lambda_i^w - \bar{\lambda}_{g(i)}^w$ is not identically zero and the floor in part (ii) persists. Time FE ($G = 1$) and every $g \times t$ specification are the same estimator applied at coarser or finer partitions: interacting dummies relocates

both failures down the hierarchy but cannot eliminate them.

(iv) **Additive two-way specification.** For any linear fixed-effects specification with annihilator M_{spec} acting on the cross-section of loadings, define the effective loadings $(M_{\text{spec}}\lambda)_i$ and $(M_{\text{spec}}\mu)_i$. Then

$$\text{plim}(\hat{\alpha}_{\text{spec}} - \alpha) = \frac{\text{Cov}((M_{\text{spec}}\mu)_i, (M_{\text{spec}}\lambda)_i) \cdot \text{Var}(\tilde{F}_t)}{\text{Var}(\check{x}_{it}^{\text{spec}})}. \quad (25)$$

For the additive two-way specification $[unit + g \times t + h \times t \text{ dummies}]$, with $N = GH$ and one unit per (g, h) -cell, refine the decomposition to $\lambda_i = \bar{\lambda} + \lambda_{g(i)}^g + \lambda_{h(i)}^h + \lambda_i^r$ (identically for μ_i). The cross-sectional annihilator of $[unit, g \times t, h \times t]$ applied to λ yields the double-demeaned residual $(M_{g,h}\lambda)_i \approx \lambda_i^r - \bar{\lambda}_{g(i)}^r - \bar{\lambda}_{h(i)}^r + \bar{\lambda}^r$, so the bias in (25) retains only the residual-layer covariance $\rho_r \sigma_{\mu,r} \sigma_{\lambda,r}$. Part (ii) applies with σ_w^2 equal to the dispersion of $(M_{g,h}\lambda)$.

The projection argument also applies when each (g, h) -cell contains multiple units, with within-cell loading heterogeneity remaining after the group effects are removed.

Proof. Parts (i)–(iii). The unit-plus- $g \times t$ within-transformation is

$$\check{z}_{it} := z_{it} - \bar{z}_{i\cdot} - \bar{z}_{g(i),t} + \bar{z}_{g(i),\cdot},$$

where $\bar{z}_{g,t} = n_g^{-1} \sum_{j: g(j)=g} z_{jt}$. Applied to $x_{it} = \mu_i F_t + \xi_{it}$:

$$\check{x}_{it} = (\mu_i - \bar{\mu}_{g(i)})(F_t - \bar{F}) + \check{\xi}_{it}, \quad (26)$$

where $\bar{\mu}_g = n_g^{-1} \sum_{j: g(j)=g} \mu_j$. Using the layer decomposition, $\mu_i - \bar{\mu}_{g(i)} = \mu_i^w - \bar{\mu}_{g(i)}^w$: the g -layer cancels exactly because $\bar{\mu} + \mu_{g(i)}^g$ appears in both μ_i and $\bar{\mu}_{g(i)}$. Identically, $\check{y}_{it} = \alpha \check{x}_{it} + (\lambda_i^w - \bar{\lambda}_{g(i)}^w) \tilde{F}_t + \check{\xi}_{it}$.

The OVB decomposition of Lemma 1 applied to the $g \times t$ -demeaned system gives

$$\text{plim}(\hat{\alpha}_{g \times t} - \alpha) = \frac{\text{Cov}(\mu_i^w - \bar{\mu}_{g(i)}^w, \lambda_i^w - \bar{\lambda}_{g(i)}^w) \cdot \text{Var}(\tilde{F}_t)}{\text{Var}(\check{x}_{it})}.$$

Computing the covariance by the same steps as Lemma 1 Step 5 with N replaced by n_g :

$$\text{Cov}(\mu_i^w - \bar{\mu}_{g(i)}^w, \lambda_i^w - \bar{\lambda}_{g(i)}^w) = \rho_w \sigma_{\mu,w} \sigma_{\lambda,w} \left(1 - \frac{1}{n_g}\right) = \rho_w \sigma_{\mu,w} \sigma_{\lambda,w} \left(1 - \frac{G}{N}\right),$$

where independence of (μ_i^w, λ_i^w) across units within the group is used, exactly as in the Lemma. Substituting into the OVB formula yields (23). At $G = 1$: $n_g = N$, $(1 - G/N) = (1 - 1/N)$, and (23) reduces to Lemma 1. This proves part (i).

For part (ii), the factor term in the score of $\hat{\alpha}_{g \times t}$ is

$$A^{(g)} = \sum_{i,t} \check{x}_{it} (\lambda_i^w - \bar{\lambda}_{g(i)}^w) \tilde{F}_t.$$

Its dominant sub-term $A_1^{(g)} = (\sum_i (\mu_i^w - \bar{\mu}_{g(i)}^w)(\lambda_i^w - \bar{\lambda}_{g(i)}^w)) S_F$ with $S_F = \sum_t \tilde{F}_t^2$ satisfies $\text{Var}(A_1^{(g)}) = O(N) \cdot O(T^2) = O(NT^2)$, because the within-group deviations are mean-zero across all N units and independent across units in different groups (and approximately independent within groups for large n_g). The denominator $D = O(NT)$, so $\text{Var}(A_1^{(g)})/D^2 = O(1/N)$. The idiosyncratic term $B^{(g)} = \sum_{i,t} \check{x}_{it} \check{\varepsilon}_{it}$ satisfies $\text{Var}(B^{(g)})/D^2 = O(1/NT)$ by the same argument as Lemma 3 Step 4. The robust sandwich estimates only $O(1/NT)$, missing the $O(1/N)$ floor, so $\hat{\text{se}}/\text{se} = O(1/\sqrt{T})$ and coverage $\rightarrow 0$ as $T \rightarrow \infty$. The floor level equals $O(\sigma_{\mu,w}^2 \sigma_{\lambda,w}^2 / N)$, that is $O(\sigma_w^4 / N)$ with $\sigma_w = \sqrt{\sigma_{\mu,w} \sigma_{\lambda,w}}$: the variance of $A_1^{(g)}$ runs through the product of the two within-group deviations, so each dispersion enters squared. The rate $O(1/\sqrt{T})$ does not depend on G .

Part (iii) is immediate: at $G = N$, $n_g = 1$ and each unit $\times$ time cell is spanned exactly by one dummy; the regression is collinear with the data. For $G < N$ at least two units share a cell, so $\lambda_i^w - \bar{\lambda}_{g(i)}^w \not\equiv 0$ and the floor persists.

Part (iv). For any FE specification, Frisch–Waugh yields $\hat{\alpha}_{\text{spec}} - \alpha = (A_{\text{spec}} + B_{\text{spec}})/D_{\text{spec}}$, where $A_{\text{spec}} = \sum_{i,t} (M_{\text{spec}} x)_{it} (M_{\text{spec}} \lambda)_i \tilde{F}_t$. The numerator is the OVB numerator of Lemma 1 applied to the residualized data; taking probability limits delivers (25). This is a simple consequence of the Frisch–Waugh theorem.

For the additive two-way case, $M_{g,h}$ acts on the loading vector by subtracting the g -group mean, the h -group mean, and adding the grand mean (the standard double-demeaning operator in the cross-section). The λ^g layer has zero g -group mean variation after subtracting $\bar{\lambda}_{g(i)}^g$, and zero h -group mean variation after the h -step (since h -group means of a g -indexed quantity equal the grand mean by the balanced structure). Identically for the λ^h layer. Hence $(M_{g,h} \lambda)_i = \lambda_i^r - \bar{\lambda}_{g(i)}^r - \bar{\lambda}_{h(i)}^r + \bar{\lambda}^r$, and the bias in (25) retains only $\text{Cov}(M_{g,h} \mu, M_{g,h} \lambda)$, which depends on $\rho_r, \sigma_{\mu,r}, \sigma_{\lambda,r}$ only. $\square$ $\square$

Remark 8 (Rank contrast and attenuation under measurement error). A $g \times t$ specification is a factor model of rank G with loadings constrained to the group indicators; IFE at rank 1

estimates one factor path with N free loadings and removes only the rank-one common component. Under classical measurement error in x this rank asymmetry bites: as the partition is refined, $g \times t$ dummies absorb an increasing fraction of the *signal* in x while the noise floor is unaffected, so attenuation compounds down the ladder even as the omitted-variable bias of part (i) shrinks – the two move in opposite directions, and the estimator can drift monotonically away from α . IFE escapes the compounding, not the attenuation: what it removes is fixed by the rank rather than by the fineness of a partition, but identifying from idiosyncratic variation – the component in which the noise share is largest – it can attenuate more than any rung of the ladder (Experiment G11, Panel C). Addressing measurement error requires an additional source of identification, i.e. an instrument; switching between the two least-squares estimators does not resolve it.

Experiment G11: group $\times$ time fixed effects. Tables A.6–A.8 evaluate Corollary 2 in a design with g -group, h -group and residual loading heterogeneity ($G = 25$, $H = 10$). The simulated biases closely match the analytical predictions: interacted fixed effects reduce bias when they absorb the dimension along which loadings covary, but residual heterogeneity continues to distort estimation and inference. Rank-one IFE remains approximately unbiased with near-nominal coverage in the absence of measurement error.

With classical measurement error, finer fixed effects reduce factor confounding but can increase attenuation by removing informative regressor variation. IFE is also subject to this limitation and exhibits the greatest attenuation in this calibration. Instrumenting with the second lag substantially reduces coefficient bias.

Experiment G12: feedback and concentration. Table A.9 examines how economic concentration affects IFE under general-equilibrium feedback. Unit 1 receives a share $s_1 \in \{1/N, 0.10, 0.25, 0.50\}$, feedback strength varies over $\psi \in \{0.5, 1.5\}$, and $\rho_{\lambda\mu} = 0$ isolates the feedback channel. As concentration and feedback increase, rank-one IFE exhibits substantial undercoverage despite limited coefficient bias: the largest absolute mean bias is 0.007, but at $(\psi, s_1) = (1.5, 0.50)$ the Monte Carlo standard deviation is nine times the reported standard error. The main distortion in this calibration therefore concerns inference.

Rank-two IFE accommodates the additional common component $\psi\theta_{ij}\bar{\xi}_t^w$ and yields coverage of 0.92–0.96 across all specifications. These results illustrate that concentration can invalidate the baseline rank-one specification without precluding identification by IFE. Increasing the rank is effective when the feedback component can be estimated and sufficient identifying variation remains after its removal.

TABLE A.6: Group×Time Fixed Effects

Estimator	Baseline calibration	Sensitivity: dominant residual layer
Time FE	0.871 {0.02}	0.899 {0.01}
<i>Corollary predicted bias</i>	−0.123	
g×t	0.969 {0.15}	0.947 {0.07}
<i>Corollary predicted bias</i>	−0.027	
h×t	0.862 {0.01}	0.892 {0.02}
<i>Corollary predicted bias</i>	−0.129	
Two-way FE	0.966 {0.08}	0.944 {0.04}
<i>Corollary predicted bias</i>	−0.028	
IFE (rank 1)	1.000 {0.95}	1.000 {0.94}
<i>Condition: implied overall $\rho_{\lambda\mu} \approx -0.285$. {Coverage}</i>		

Notes: 500 replications. $N = 250$ ($G = 25 \times H = 10$), $T = 100$, $\alpha = 1$, $\rho_F = 0.8$, $\rho_\xi = 0.8$. Baseline layer calibration: $\sigma_g^2 = 0.6$, $\rho_g = -0.4$; $\sigma_h^2 = 0.1$, $\rho_h = 0$; $\sigma_r^2 = 0.3$, $\rho_r = -0.15$. Sensitivity: $\sigma_g^2 = 0.3$, $\sigma_r^2 = 0.6$ (layers permuted). Predicted bias from Corollary 2 via the projection principle (per-replication average).

TABLE A.7: Group×Time Fixed Effects: Coverage Collapse in T

Estimator	<i>Time periods T</i>		
	$T = 50$	$T = 100$	$T = 200$
Time FE	0.869 {0.03}	0.865 {0.01}	0.864 {0.02}
Two-way FE	0.966 {0.16}	0.966 {0.09}	0.967 {0.06}
IFE (rank 1)	1.000 {0.96}	1.000 {0.96}	1.000 {0.94}
<i>Condition: baseline layer calibration as in Panel A. {Coverage}</i>			

Notes: 500 replications. $N = 250$. Coverage of Time FE and Two-way FE declines in T at the same qualitative rate (Corollary 2(ii) rate-invariance); IFE coverage is flat at the nominal level.

TABLE A.8: Group×Time Fixed Effects: Attenuation under Measurement Error

Estimator	Estimate
Time FE	0.735 {0.00}
g×t	0.631 {0.00}
h×t	0.732 {0.00}
Two-way FE	0.614 {0.00}
IFE (rank 1)	0.566 {0.00}
Two-way + IV	0.976 {0.60}
<i>Condition: classical measurement error in x. {Coverage}</i>	

Notes: 500 replications. $N = 250$, $T = 100$; μ g -layer scaled by $5\times$, $\sigma_\xi = 0.7$. Measurement error: $x_{\text{obs}} = x + u$, $u \sim N(0, \sigma_{\text{me}}^2)$, $\sigma_{\text{me}} = 1.38$; realized full-sample reliability ≈ 0.86 , realized g×t share of $\text{Var}(x) \approx 0.70$. IV with a second-lag instrument.

Experiment G13: measurement error under a shared factor. Table A.10 repeats Experiment G6 when the outcome and regressor share a common factor. Removing this factor leaves IFE with less identifying signal relative to measurement noise than Time FE. Consequently, IFE exhibits greater attenuation at every positive noise share considered. The comparison confirms that the relative sensitivity of the two estimators to measurement error depends on the factor structure; the threshold reported in Experiment G6 is specific to that calibration.

Experiment G14: mean-group and common correlated effects estimators. This experiment confronts IFE with the two estimators the heterogeneous-panel literature offers for the same problem. The mean-group (MG) estimator of Pesaran and Smith (1995) runs the regression unit by unit and averages the slopes. The common correlated effects (CCE) estimator of Pesaran (2006) first augments each unit’s regression with the cross-sectional averages $(\bar{x}_t, \bar{y}_t)$, which span the factor space when the matrix of loading means has full rank, and then either pools the slope (CCE-P) or averages it (CCE-MG). Both are defactoring devices that use the averages as observed proxies for the factors rather than estimating the factors, and neither requires a choice of rank. The design is that of Section 5 with one addition: the loadings share a mean $\bar{\mu}$, so that $\lambda_i = \bar{\mu} + \sigma z_{1i}$ and $\mu_i = \bar{\mu} + \sigma(\rho_{\lambda\mu} z_{1i} + \sqrt{1 - \rho_{\lambda\mu}^2} z_{2i})$. That parameter governs CCE’s rank condition: at $\bar{\mu} = 0$ the averages are

TABLE A.9: Monte Carlo: GE Feedback and Concentration

ψ	Estimator	Dominant-unit share s_1 ($\max_i s_i$)			
		$1/N$	0.10	0.25	0.50
0.5	Time FE	1.001 {0.31}	1.002 {0.32}	1.001 {0.32}	1.000 {0.34}
	IFE (rank 1)	1.001 {0.93}	1.001 {0.95}	0.999 {0.90}	0.995 {0.64}
	IFE (rank 2)	1.000 {0.95}	1.000 {0.96}	1.000 {0.94}	1.000 {0.95}
1.5	Time FE	1.004 {0.31}	1.002 {0.32}	0.999 {0.31}	0.997 {0.28}
	IFE (rank 1)	1.001 {0.91}	1.000 {0.90}	0.993 {0.57}	0.999 {0.33}
	IFE (rank 2)	1.000 {0.94}	1.000 {0.96}	1.000 {0.96}	1.000 {0.92}
<i>Condition: $\rho_{\lambda\mu} = 0$ (feedback channel isolated). {Coverage}</i>					

Notes: 500 replications. Baseline calibration with GE feedback $\psi\theta_{ij}\bar{X}_t^w$ added to the outcome, where $\bar{X}_t^w = \sum_{ij} s_{ij}x_{ijt}$, $\theta_{ij} \sim \mathcal{N}(1, 1)$, unit 1 carries share s_1 and the remaining $N - 1$ units share the rest equally, so $\text{HHI} \approx s_1^2$ at high concentration and $1/N$ at equal weights. $\rho_{\lambda\mu} = 0$ isolates the feedback channel: all bias here is GE contamination.

TABLE A.10: Measurement Error in the Regressor: Shared Factor

	<i>Noise share of the observed regressor</i>					
	0	0.10	0.25	0.50	0.75	0.90
Attenuation benchmark	1.000	0.900	0.750	0.500	0.250	0.100
Time FE	0.998 {0.35}	0.851 {0.00}	0.657 {0.00}	0.391 {0.00}	0.175 {0.00}	0.066 {0.00}
IFE (rank 1)	0.999 {0.93}	0.749 {0.00}	0.505 {0.00}	0.256 {0.00}	0.102 {0.00}	0.037 {0.00}
Time FE – IFE	–0.002	0.102	0.151	0.135	0.073	0.029
<i>Share of the observed regressor surviving each transformation</i>						
Time FE	0.668	0.705	0.756	0.839	0.918	0.965
IFE (rank 1)	0.331	0.399	0.506	0.671	0.832	0.927

Notes: 500 replications. Same design as Experiments G1–G2: a single common factor in both equations, $y_{it} = \alpha x_{it} + \lambda_i F_t + \varepsilon_{it}$ and $x_{it} = \mu_i F_t + \xi_{it}$, with the regressor observed with classical error, $x_{it}^{\text{obs}} = x_{it} + \nu_{it}$, and the noise share defined as $\text{Var}(\nu)/\text{Var}(x^{\text{obs}})$. $\alpha = 1$, $N = 250$, $T = 100$, $\rho_F = \rho_\xi = 0.80$, $\lambda_i, \mu_i \sim \mathcal{N}(1, 1)$ independent ($\rho_{\lambda\mu} = 0$, omitted-variable channel shut). Coverage of the true $\alpha = 1$ by nominal 95% confidence intervals, in braces. The attenuation benchmark is $\alpha \cdot \text{Var}(x)/\text{Var}(x^{\text{obs}})$. The bottom panel reports the share of the within variance of the observed regressor surviving each estimator’s transformation.

$O(N^{-1/2})$ noise, at $\bar{\mu} = 0.5$ they are informative proxies. Standard errors are the cross-sectional dispersion of the unit slopes for MG and CCE-MG and a sandwich over units for CCE-P; Time FE appears twice, with heteroskedasticity-robust and with unit-clustered standard errors (Appendix B.4).

Table A.11 compares MG, CCE and IFE across the three experiments. MG does not remove factor confounding; averaging unit-specific slopes retains their omitted-variable bias. It can therefore be biased even when Time FE is approximately unbiased. CCE performs comparably to IFE when $\bar{\mu} = 0.5$, because the cross-sectional averages capture the common factor. When $\bar{\mu} = 0$, those averages are uninformative, and CCE retains 39–52 percent of the Time FE bias under correlated loadings.

Panel B allows the synchronized and idiosyncratic slopes to differ. With informative averages, CCE removes the synchronized component and recovers α_x . With zero-mean loadings, its estimates lie between the two component slopes. IFE recovers α_x in both cases because it estimates the shared factor directly rather than approximating it through cross-sectional averages.

Table A.12 adds a second factor. Panel A uses the general-equilibrium block of Experi-

ment G7 with $\bar{\mu} = 0.5$ and $\rho_{\lambda\mu} = 0.5$: with two factors and two averages, CCE’s rank condition holds with equality. The point estimates survive (bias 0.008), but coverage falls to 0.58–0.63, with or without feedback. Panel B shows why. The intervals are close to correctly sized, except at the smallest N – evaluated at the estimator’s own Monte Carlo mean they cover at 0.84–0.94 – and the loss is a bias too large for intervals that tight to absorb. That bias is the finite- N error in the proxies and it falls like $1/N$; but it does not depend on T , whereas the intervals shrink with T , so lengthening the panel from 50 to 200 periods lowers coverage from 0.75 to 0.34. Calibration returns only as $T/N \rightarrow 0$, a limit macro panels do not approach. Two further cells isolate the mechanism. The degradation needs the exhausted rank condition and correlated loadings together – either alone leaves coverage near nominal (not tabulated) – and when the loadings have mean zero while the second factor is mean-loaded, the averages proxy the wrong factor and CCE carries 55–76 percent of the Time FE bias (last column of Panel A). IFE at rank two is centred with near-nominal coverage in every cell; at rank one it fails, as in Experiment G7 – here through the point estimate, because the loadings are correlated. The comparison turns on where each estimator’s requirement lives. IFE’s is a single stated one, the rank. CCE asks less of the researcher – no rank has to be chosen – and where its conditions hold it matches IFE on the one slope it can identify. But those conditions are statements about loading means and about a proxy budget, on neither of which the estimator reports, and each failure above – the blend at $\bar{\mu} = 0$, the wrong factor, the miscalibrated intervals at the boundary – is silent from inside the estimator.

Experiment G15: one aggregate or two. Step 4 of the checklist compares the regressor’s factor with the outcome’s. Table A.13 generates data from (4) with independent loadings and varies the correlation between F_t^x and F_t^y from zero to one, for three synchronized shares of the within variance of the regressor, panel sizes $N, T \in \{50, 100, 250\}$, two degrees of persistence and two values of $\alpha_c - \alpha_x$. The first two columns report the correlation between the first principal component of the regressor and the factor of the pooled IFE regression above which the pooled slope is within ten percent of $\alpha_c - \alpha_x$ from α_x , so that a single regression suffices. The threshold rises with the synchronized share and varies little with the panel size and with persistence. When $\alpha_c - \alpha_x$ is large the pooled factor rotates toward F_t^x , so this correlation overstates the one between the two aggregates; the factor of the two-step regression does not rotate. The remaining columns report the two-step estimate of α_c : its bias and dispersion grow with the correlation between the aggregates, moderately up to 0.8 and quickly beyond 0.9, and α_c is not identified when the two coincide. The two-step estimate of α_x is within about one percent of $\alpha_c - \alpha_x$ from its true value in every cell.

TABLE A.11: Mean-Group and Common-Correlated-Effects Estimators: The Three Failures

<i>Panel A: Failure I, $T = 100$</i>								
Estimator	$\bar{\mu} = 0$				$\bar{\mu} = 0.5$			
	$\rho_{\lambda\mu} = 0$		$\rho_{\lambda\mu} = 0.5$		$\rho_{\lambda\mu} = 0$		$\rho_{\lambda\mu} = 0.5$	
	Bias	Cov.	Bias	Cov.	Bias	Cov.	Bias	Cov.
Time FE	−0.002	0.26	0.246	0.00	−0.001	0.31	0.243	0.00
Time FE, clustered by unit	−0.002	0.93	0.246	0.00	−0.001	0.95	0.243	0.00
MG	−0.002	0.94	0.173	0.00	0.084	0.20	0.230	0.00
CCE-P	−0.001	0.94	0.127	0.00	0.000	0.96	0.002	0.89
CCE-MG	−0.001	0.95	0.096	0.00	0.000	0.97	0.003	0.91
IFE (rank 1)	0.000	0.94	0.000	0.95	0.000	0.96	0.000	0.94
<i>Panel B: Failure II, $\alpha_c = 2$, $\alpha_x = 1$</i>								
Estimator	$\bar{\mu} = 0$		$\bar{\mu} = 0.5$					
	Mean	s.d.	Mean	s.d.				
Time FE	1.494	0.084	1.485	0.083				
MG	1.348	0.056	1.461	0.065				
CCE-P	1.218	0.141	1.002	0.004				
CCE-MG	1.169	0.095	1.002	0.004				
IFE (rank 1)	1.000	0.004	1.000	0.004				
<i>Panel C: Failure III, $\rho_{\lambda\mu} = 0$. Coverage</i>								
Estimator	$\bar{\mu} = 0$			$\bar{\mu} = 0.5$				
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$		
Time FE	0.41	0.26	0.24	0.46	0.31	0.21		
Time FE, clustered by unit	0.94	0.93	0.94	0.96	0.95	0.95		
MG	0.95	0.94	0.95	0.23	0.20	0.19		
CCE-P	0.94	0.94	0.94	0.96	0.96	0.97		
CCE-MG	0.95	0.95	0.95	0.95	0.97	0.96		
IFE (rank 1)	0.94	0.94	0.95	0.96	0.96	0.97		

Notes: 500 replications. DGP: $\alpha = 1$, $\rho_F = \rho_\xi = 0.80$, $N = 250$, $\sigma = 1$; $\bar{\mu}$ is the mean loading. Bias is the mean estimate minus 1; Cov. is coverage. Time FE uses heteroskedasticity-robust standard errors unless clustered by unit. MG averages unit-by-unit slopes; CCE augments each unit's regression with $(\bar{x}_t, \bar{y}_t)$ (Pesaran, 2006). Panel B: $\alpha_x = 1$, $\alpha_c = 2$; the Time FE estimate does not depend on the standard errors, so one row is shown.

TABLE A.12: Common-Correlated-Effects Estimators at the Boundary of the Rank Condition

<i>Panel A: two factors, $\rho_{\lambda\mu} = 0.5$, $N = 250$, $T = 100$</i>						
Estimator	$\psi = 0$		$\psi = 0.5$		$\psi = 0.5, \bar{\lambda} = \bar{\mu} = 0$	
	Bias	Coverage	Bias	Coverage	Bias	Coverage
Time FE	0.245	0.00	0.247	0.00	0.245	0.00
MG	0.235	0.00	0.341	0.00	0.177	0.00
CCE-P	0.008	0.59	0.008	0.58	0.187	0.00
CCE-MG	0.008	0.63	0.008	0.62	0.135	0.00
IFE (rank 1)	0.221	0.01	0.170	0.00	0.226	0.00
IFE (rank 2)	0.000	0.95	0.000	0.94	0.000	0.95

<i>Panel B: CCE-P at the boundary ($\psi = 0$) as T and N vary</i>					
	Bias	s.d. of $\hat{\alpha}$	Median s.e.	Coverage	Recentred coverage
$N = 250$					
$T = 50$	0.0091	0.0080	0.0069	0.75	0.93
$T = 100$	0.0079	0.0055	0.0046	0.62	0.91
$T = 200$	0.0081	0.0041	0.0033	0.34	0.90
$T = 100$					
$N = 100$	0.0205	0.0141	0.0087	0.45	0.84
$N = 250$	0.0079	0.0055	0.0046	0.62	0.91
$N = 500$	0.0042	0.0034	0.0031	0.73	0.92
$N = 1000$	0.0021	0.0022	0.0021	0.83	0.94

Notes: The true slope is 1. Panel A: 500 replications, $\bar{\mu} = 0.5$; a second factor $M_t = 0.8M_{t-1} + \psi\bar{x}_t + u_t^M$ enters the outcome with loadings $\theta_i \sim \mathcal{N}(1, 1)$; in the last pair of columns (λ_i, μ_i) have mean zero. Time FE coverage uses heteroskedasticity-robust standard errors and is also 0.00 in every column when they are clustered by unit. Panel B: 500 replications. Recentred coverage evaluates the same intervals at the estimator's Monte Carlo mean.

TABLE A.13: One aggregate or two

Sync share	Threshold, pooled IFE		Two-step α_c , by correlation between F_t^x and F_t^y						
	small gap	large gap	0	0.3	0.6	0.8	0.9	0.95	0.99
30 percent	0.79–0.81	0.72–0.77	−7.0 [1.00]	−7.8 [1.08]	−10.0 [1.26]	−14.0 [1.48]	−20.0 [1.97]	−30.8 [2.47]	−68.2 [3.88]
50 percent	0.91	0.87–0.89	−3.0 [1.00]	−3.8 [1.04]	−4.7 [1.16]	−7.0 [1.56]	−10.1 [2.07]	−17.0 [2.81]	−49.6 [5.32]
70 percent	0.96	0.93–0.94	−1.4 [1.00]	−1.7 [0.96]	−2.4 [1.14]	−3.5 [1.50]	−5.7 [2.10]	−9.0 [2.95]	−34.1 [7.67]

Notes: Synchronized share: fraction of within regressor variance explained by $\mu_i F_t^x$. Threshold: correlation between the regressor’s first principal component and the pooled IFE factor above which the pooled slope differs from α_x by less than $0.1|\alpha_c - \alpha_x|$. Ranges cover persistence $\{0, 0.5\}$ at $N = T = 100$ for the small gap (0.15), and additionally $N, T \in \{50, 100, 250\}$ for the large gap (1.5). Two-step α_c : median bias as a percentage of $\alpha_c - \alpha_x$; brackets report the Monte Carlo standard deviation relative to that at zero factor correlation. $R = 500$.

D Complements to the Application

D.1 Detailed Data Sources and Descriptions

In this section we describe the sources of our data and any transformation applied.

Our primary data source is the Joint Harmonised EU Programme of Business and Consumer Surveys, administered by the Directorate-General for Economic and Financial Affairs of the European Commission⁴³.

These surveys are conducted at the national level by various EU Commission’s partner institutes, including ministries, statistical offices, central banks, research institutes, business associations and private companies. The program follows a harmonized methodology to ensure comparability across countries.

The survey samples are taken out from national frames intended to capture the entire relevant population. These frames are derived from official or statistical registers, as well as membership lists from business associations and chambers of commerce. Given the dynamic nature of business populations – due to mergers, bankruptcies and the entry of new firms – the sample frames are regularly updated to maintain representativeness.

⁴³The following description is based on the official methodological user guide of the Commission, available here [bcs_user_guide.pdf](#) – for a more detailed overview.

Survey samples are designed to ensure efficiency and precision. Many national institutes employ stratified random sampling, which divides the population into non-overlapping subgroups (strata) with similar variance concerning key variables. Business survey stratification is typically based on firm size and sector of activity.

The Business and Consumer Surveys operate under two fundamental principles to achieve harmonization across countries:

- all national institutes use the same harmonized questionnaires;
- surveys are conducted and results transmitted according to a common timetable.

Survey timing is consistent across countries. Questionnaires are administered in the first two to three weeks of each month for the monthly surveys and quarterly ones are carried out during the first month⁴⁴ of each quarter (January, April, July, and October).

The panel structure of business surveys ensures consistency over time. In most cases, companies included in the sample remain in the survey for multiple periods, with periodic updates or partial replacements (rotating panels). This approach improves response speed, reduces variability between successive surveys and optimizes sample size.

We apply the following data-cleaning steps to address missing values and outliers. First, we focus on those series for which we have at least (i) half of the sample time frame available and (ii) that are also different from zero⁴⁵. This ensures that our sample includes countries and sub-sectors for which we have a reasonable amount of data and that experienced hiring constraints. Second, we exclude values below zero and above one-hundred, as these are clearly reporting issues, as these variables are in percentage terms and cannot be negative or above one-hundred.

Due to the relatively high correlation between PLS and supply shortages, before estimating the main results we decided to orthogonalize supply shortages having it regressed on PLS and using the residuals in the regressions. This ensures that only the variation in supply shortages, independent of PLS, is used. The choice affects what the PLS coefficient in the controls specifications means, and the interpretation should be stated. Because only the component of supply shortages orthogonal to PLS enters, the PLS coefficient retains whatever inflation-relevant variation the two measures share; it is not a coefficient partialled out with respect to the raw supply-shortage series. The estimand is thus the effect of labor shortages inclusive of their common component with reported supply bottlenecks, which

⁴⁴Always around the first two to three weeks.

⁴⁵The main results are robust to this choice.

is the object we want given that the two survey items plausibly reflect a single underlying capacity constraint – but a specification controlling for raw supply shortages would attribute that shared variation to the control instead and would deliver a smaller PLS coefficient. We use capacity constraint as an alternative measure to supply shortages and confirm the main results.

The second set of variables we employ come from different sources. The producer price index is downloaded directly from Eurostat at quarterly frequency both in percentage change compared to same period in previous year and as an index. The number of vacancies and unemployed people are also downloaded from Eurostat. In order to adjust for the difference in the measurement scale of the two variables, we divide vacancies by one-thousand. Once the ratio is computed, we apply the natural logarithm transformation. Energy prices come also from the IMF and are available through their commodity data portal, while the nominal effective exchange rate is downloaded from the BIS data portal.

For the extended validation exercise reported in Table 7, we additionally download three series from Eurostat. The unemployment rate (age 20–64, percentage of active population, seasonally adjusted) is taken from Eurostat’s labor force survey; the complement $1 - u$ is computed as one minus the unemployment rate. Unit labor costs are downloaded from Eurostat’s national accounts: nominal unit labour costs per hour worked, not seasonally adjusted, expressed as percentage change compared to the same quarter of the previous year. The labor share is constructed as the ratio of compensation of employees to gross domestic product at current market prices, both downloaded from Eurostat’s quarterly national accounts. All three series are available at quarterly frequency and cover the same sample as the main analysis.

Variables for the U.S. come from three sources. The first one is the National Federation of Independent Business’ *Small Business Economic Trends* indicators, which are publicly available. We downloaded and used the series with current job openings, which gives the percent of respondents with a position they are currently not able to fill (at the national level). This measure is indeed close to the PLS one we employed from the EU Commission’s Business and Consumer Surveys. Regarding the vacancies and unemployment data for the U.S. (necessary to compute the V/U ratio), we use the same data as in [Benigno and Eggertsson \(2023\)](#). We thank the authors for kindly sharing their data.

The second source is the Federal Reserve Bank of St. Louis’ FRED database. We download three series. The unemployment rate, seasonally adjusted, is aggregated from monthly to quarterly frequency by taking the within-quarter average; the complement $1 - u$ is then com-

puted as one minus the unemployment rate. Unit labor costs for the nonfarm business sector are available at quarterly frequency as an index; we compute the year-on-year percentage change to obtain a growth rate comparable to the euro area series. The nonfarm business labor share index is used directly as a measure of the labor share.

D.2 List of Manufacturing Sub-Sectors

TABLE A.14: Manufacturing Sub-Sectors

Sector definition	NACE 2-digit Codes	Sub-sectors
Food products, beverages, and tobacco products	10 – 12	Food products, Beverages, Tobacco products
Textiles, apparel, leather, and related products	13 – 15	Textiles, Wearing apparel, Leather and related products
Wood and paper products, and printing	16 – 18	Wood products, Paper and paper products, Printing and recorded media
Chemicals and chemical products	20	Chemicals and chemical products
Rubber and plastics products, and other non-metallic mineral products	22 – 23	Rubber and plastic products, Other non-metallic mineral products
Basic metals and fabricated metal products	24 – 25	Basic metals, Fabricated metal products
Computer, electronic, and optical products	26	Computer, electronic and optical products
Electrical equipment	27	Electrical equipment
Machinery and equipment n.e.c.	28	Machinery and equipment n.e.c.
Transport equipment	29 – 30	Motor vehicles, trailers and semi-trailers, Other transport equipment
Other	19, 21, 31, 32, 33	Manufacture of coke and refined petroleum products, Manufacture of basic pharmaceutical products and pharmaceutical preparations, Manufacture of furniture, Other manufacturing, Repair and installation of machinery and equipment

D.3 Summary Statistics

[↔ Back to reference in main text](#)

Table A.15 presents summary statistics for the key variables over the full sample period, as well as separately for the pre-COVID and post-COVID periods, given the scale of the pandemic shock.

TABLE A.15: Summary Statistics

	Full Sample		Pre-Covid		Post-Covid	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Inflation	1.483	2.120	0.891	0.915	3.760	3.533
PLS	8.369	6.414	5.954	3.476	17.649	6.715
Supply Shortages	11.353	11.760	7.045	3.421	27.905	16.968
Insufficient Demand	29.967	9.049	30.584	8.829	27.598	9.733
Financial Constraints	3.893	1.094	3.989	1.166	3.524	0.656
WEO Inflation Expectations	1.815	0.158	1.805	0.171	1.856	0.070
Energy Prices	-0.000	57.086	-0.869	43.937	3.337	93.480
NEER	-0.000	0.108	0.001	0.099	-0.004	0.140
Other Constraints	7.148	3.925	5.860	0.986	12.098	6.434
Wage Inflation	2.526	1.225	2.394	0.817	3.035	2.138

Notes: The table reports means and standard deviations for the full sample (2002Q1–2024Q4), the pre-Covid period (2002Q1–2019Q4) and the post-Covid period (2020Q1–2024Q4). The two sub-periods are non-overlapping and together exhaust the full sample. Aggregates are weighted averages based on official EU Commission weights.

D.4 Estimates of the Common Component (Rank Selection)

[↔ Back to reference in main text](#)

Across the methods considered and detailed in Online Appendix OA.1.3, results in Table A.16 uniformly support a single common component (F_t^y , in our baseline equation): the six panel information criteria of Bai and Ng (2002), the eigenvalue-ratio and growth-ratio criteria of Ahn and Horenstein (2013), the diagnostic criterion of Gagliardini et al. (2019) and the Edge Distribution estimator of Onatski (2010) all select one factor, in both specifications and for a maximum of ten factors considered. Two features of this result are worth emphasizing. First, applied to the fixed-effects residuals, all criteria select a single common factor, supporting the baseline rank-one specification. Second, the single estimated factor carries a clear economic signature (Figure A.1), consistent with the literature that emphasizes inflation co-movements

and suggests that one factor alone explains a large proportion of cross-country inflation variation (Ciccarelli and Mojon, 2010; Auer et al., 2025). This result lends empirical support to the view that inflation is shaped by common shocks, which we interpret as potentially related to (i) shared inflation expectations dynamics within the monetary union and (ii) aggregate cost-push shocks (e.g. energy shocks).

TABLE A.16: Number of Common Factors: Information Criteria

Specification	Bai and Ng (2002)		Eigenvalue-based			
	$PC_{p1,p3}$	$IC_{p1,p3}$	ER	GR	GOS	ED
PLS only	1	1	1	1	1	1
PLS and controls	1	1	1	1	1	1

Notes: Each cell reports the number of common factors selected by each criterion, with a maximum of ten factors considered. $PC_{p1}-IC_{p3}$ = panel information criteria (Bai and Ng, 2002); ER = Eigenvalue Ratio, GR = Growth Ratio (Ahn and Horenstein, 2013); GOS = Gagliardini et al. (2019); ED = Edge Distribution estimator (Onatski, 2010).

TABLE A.17: Number of Factors in PLS: Information Criteria

Specification	Bai and Ng (2002)		Eigenvalue-based			
	$PC_{p1,p3}$	$IC_{p1,p3}$	ER	GR	GOS	ED
PLS, raw	10	10	1	1	1	1
PLS, pair-demeaned	10	10	1	1	1	1

Notes: Each cell reports the number of common factors selected by each criterion, with a maximum of ten factors considered. $PC_{p1}-IC_{p3}$ = panel information criteria (Bai and Ng, 2002); ER = Eigenvalue Ratio, GR = Growth Ratio (Ahn and Horenstein, 2013); GOS = Gagliardini et al. (2019); ED = Edge Distribution estimator (Onatski, 2010).

On the regressor panel itself (Table A.17), the picture is less clear-cut. The eigenvalue-ratio and growth-ratio rules choose a single factor, whereas the Bai–Ng information criteria hit the imposed cap of ten factors. This discrepancy likely reflect the tendency of the Bai–Ng criteria to overselect under residual cross-sectional dependence. Daniele (2024) shows that the Bai and Ng (2002) criteria tend to overestimate the number of factors systematically, because remaining cross-sectional dependence in the residuals is mistaken for a common component and the penalty does not sufficiently penalize overparameterization in finite samples.

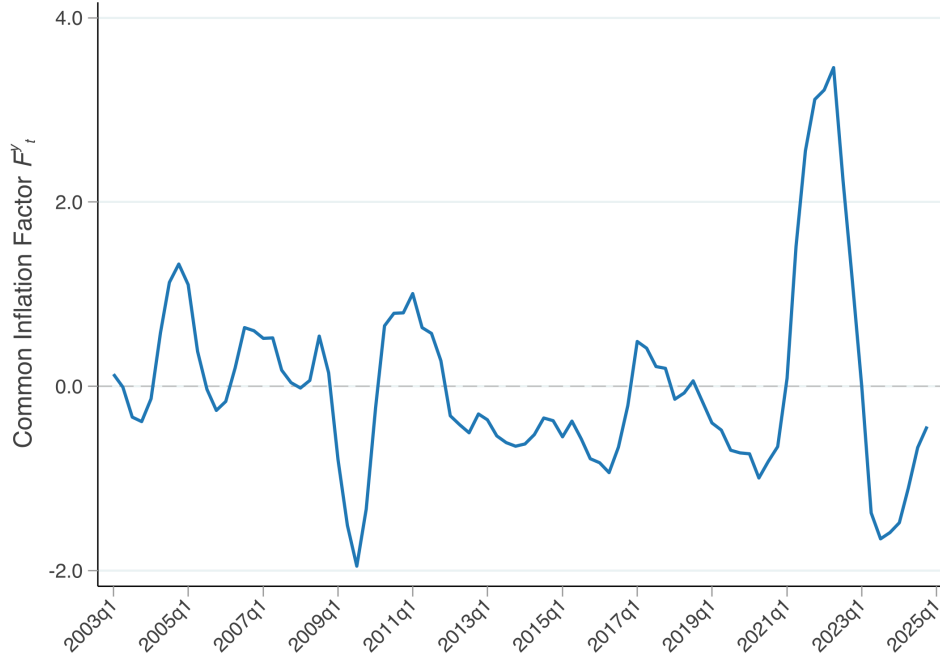
D.5 Estimated Common Factor and Distribution of Loadings

↔ *Back to reference in main text*

This section plots the outcome side of the factor structure: the estimated common factor of inflation, F_t^y , and the cross-sectional distribution of its loadings λ_{ij} . Both are taken from the featured specification, column (8) of Table 9 – that is, from the regression in which measured tightness already enters as its two components, $\mu_{ij}F_t^x$ and ξ_{ijt} , and the expectations proxy is included. The notation is the two-aggregate notation of Section 3.2: F_t^y is the aggregate the outcome loads on, F_t^x the aggregate the regressor loads on, and the figures that follow this subsection display the latter pair, F_t^x and μ_{ij} , on the same conventions.

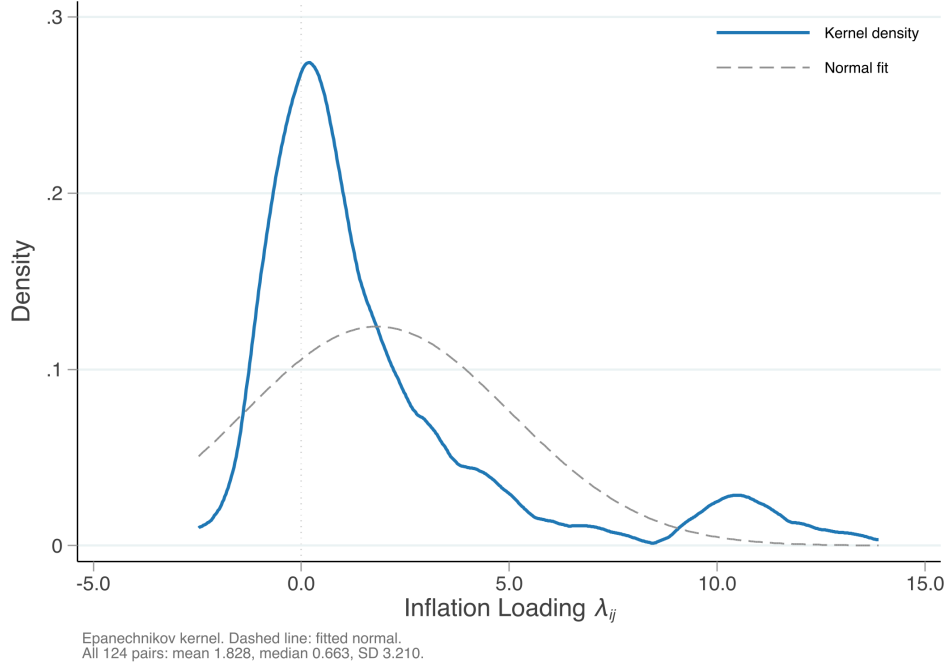
A sign convention is needed before either object can be described. In a one-factor model (F_t^y, λ_{ij}) and $(-F_t^y, -\lambda_{ij})$ are observationally equivalent. Therefore, we normalize the sign so that the cross-sectional mean of the estimated loadings is positive, $N^{-1} \sum_{ij} \hat{\lambda}_{ij} > 0$, which orients the factor to move with inflation for the typical unit; the same principle orients F_t^x so that an increase denotes a tighter labor market. All signed descriptions below, and the right-skewness of the loading distribution, are stated under that normalization. The slope α is invariant to it.

FIGURE A.1: The Common Component of Inflation, F_t^y



Notes: The figure plots the time series of the single common factor estimated jointly with the slope in the featured specification, column (8) of Table 9. The factor troughs around the Great Financial Crisis and peaks sharply during the post-pandemic inflation surge, consistent with its interpretation as capturing common cost-push and expectational shocks. The factor is identified up to scale and sign; the scale normalization is $F'F/T = I$, standard in the principal components literature, and the sign is fixed as described above.

FIGURE A.2: Cross-Sectional Distribution of the Inflation Loadings λ_{ij}



Notes: The figure plots the kernel density of the estimated loadings λ_{ij} of inflation on the common inflation factor across the sub-sector-country pairs, from the featured specification, together with a fitted normal. The distribution is markedly right-skewed and visibly departs from normality, indicating substantial heterogeneity in units' exposure to the common factor – the heterogeneity that time fixed effects suppress and that drives the attenuation documented in Appendix C.

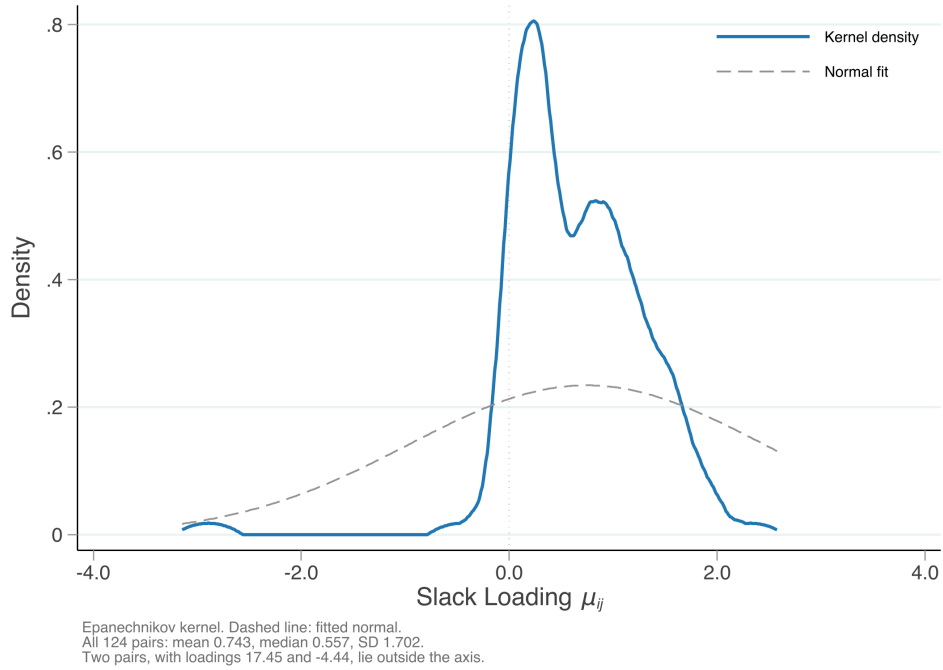
The regressor's factor structure. The two figures above describe the factor structure of the inflation equation: the common factor F_t^y that the interactive effects estimate from the outcome, and the dispersion of the loadings λ_{ij} on it. The figures that follow describe the other side of the two-aggregate case, the factor structure of the *regressor*. Where the two aggregates differ, as Lemma 2 and Remark 2 set out, recovering the micro-local slope requires defactoring PLS itself: extracting its common driver F_t^x , fitting a loading μ_{ij} for every sub-sector-country pair, and splitting measured tightness into a synchronized component $\mu_{ij}F_t^x$ and an idiosyncratic remainder. Both components then enter the estimating equation and carry their own slopes, α_c and α_x (Table 9). The four figures below display that decomposition and the heterogeneity in μ_{ij} that makes it bite.

FIGURE A.3: The Common Component of Labor Market Tightness, F_t^x



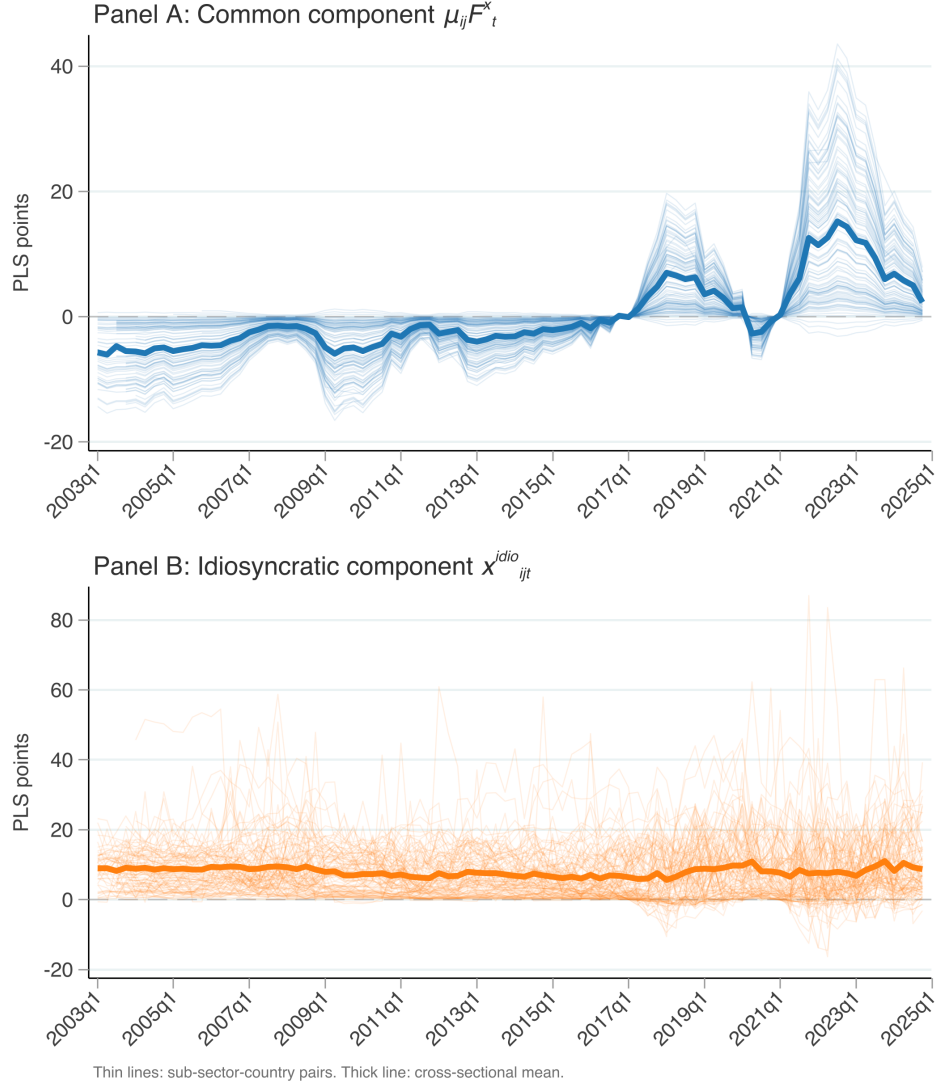
Notes: The figure plots the first principal component over time of pair-demeaned PLS. This is the object Step 5 of the checklist defactors out of the regressor before the micro-local slope is read off. The component is identified up to scale and sign; we normalize the sign so that F_t^x comoves positively with the cross-sectional mean of pair-demeaned PLS, i.e., so that an increase denotes a tighter labor market.

FIGURE A.4: Cross-Sectional Distribution of the Slack Loadings μ_{ij}



Notes: The figure plots the kernel density of the estimated loadings μ_{ij} of PLS on the common slack factor across the sub-sector-country pairs, together with a fitted normal. The distribution is markedly right-skewed and visibly departs from normality, indicating substantial heterogeneity in units' exposure to the common component of slack – the heterogeneity that a period dummy, which imposes a single exposure on every unit, cannot represent. The horizontal axis is trimmed to the bulk of the distribution: two pairs observed for only eight quarters carry loadings far outside it, and both the fitted normal and the reported standard deviation include them.

FIGURE A.5: Common and Idiosyncratic Components of PLS, All Pairs



Notes: Panel A plots the common component $\mu_{ij}F_t^x$ and Panel B the idiosyncratic remainder $x_{ijt}^{idio} = PLS_{ijt} - \mu_{ij}F_t^x$, one thin line per pair, with the cross-sectional mean in bold. Pairs observed for fewer than 40 quarters are omitted from the display, whose axes two very short pairs would otherwise set; the estimation sample is unchanged. The contrast is the identification argument in visual form: in Panel A every pair traces the same path scaled by its own loading, so the paths fan out as the factor moves and a single period dummy cannot absorb them; in Panel B there is no shared shape left to absorb.

D.6 Complements to the Estimand Decomposition

[↔ Back to reference in main text](#)

Differential attenuation. Measurement error may attenuate the idiosyncratic slope more strongly than the synchronized slope, increasing the estimated ratio. Two simulation exercises assess this possibility. Applying the two-step procedure to synthetic outcomes with equal component slopes produces no systematic difference between the estimates. Adding classical measurement error to the regressor increases the estimated ratio from unity to 1.3, 1.6 and 2.7 at noise shares of 10, 25 and 50 percent of within-pair variance, respectively. In this calibration, reproducing a ratio close to the observed 2.8 requires substantial measurement error.

Interpreting the synchronized slope. The larger synchronized coefficient may reflect a stronger inflation response to widespread shortages, but it may also capture aggregate shocks correlated with tightness. The component decomposition alone does not distinguish these explanations. Under the maintained assumptions, the idiosyncratic slope remains identified from unit-specific variation; a causal interpretation of the synchronized slope requires an additional exclusion restriction. The following IV exercise addresses this distinction.

Instrumenting the synchronized component. Remark 2 makes the causal reading of α_c conditional on an aggregate instrument. Two candidates were taken to the data. *High-frequency monetary policy surprises.* Meeting-day changes in OIS rates from the euro area monetary policy event-study database (Altavilla et al., 2019), summed by quarter over 2003–2024, load *positively* on the PLS common factor at every horizon up to ten quarters: a surprise tightening coincides with a tighter labor market, the signature of the information effect (Jarociński and Karadi, 2020). Splitting the surprises by the sign of the stock market reaction confirms it. The information component carries all the relevance (first-stage F of 24 on the eight-quarter cumulation) and moves inflation directly holding the factor fixed, while the policy component has the expected negative sign but a first-stage F between 0.3 and 1.5. The policy component has insufficient first-stage relevance, while the information component raises concerns about the exclusion restriction. *An internal instrument.* Following Sarto (2025), we write reported shortages as $PLS_{ijt} = \theta_{ij}\hat{\Omega}_{ijt} + \gamma_{ij}u_t + e_{ijt}$, where $\hat{\Omega}_{ijt}$ is pair ij 's fitted common component of inflation at the paper's rank of one, and recover u_t as the first principal component of the residual by alternating least squares over $(\theta_{ij}, \gamma_{ij}, u_t)$. The estimated θ_{ij} are small (mean 0.2), so feedback from inflation to reported shortages is not what separates the projection from a causal slope, and u_t is close to the PLS factor itself.

The instrument is the innovation to u_t interacted with the pair’s exposure $\hat{\mu}_{ij}$; it is not a persistent level, so the date-clustered interval is appropriate (Experiment G5), and weak-instrument-robust sets are reported (Anderson and Rubin, 1949). Identification rests on the restriction that this innovation is orthogonal to the common shocks that enter inflation directly; it cannot be tested from the panel. Table A.18 reports the results. In the featured specification without the estimated outcome factor – the form Sarto’s estimator takes, since an instrument orthogonal to the common shocks needs no control for them – the IV estimate is 0.19 with a first-stage F of 10.9 and a 95 percent set of $[0.06, 0.36]$, against a projection of 0.124. Holding the estimated outcome factor fixed removes part of the instrumented variation, because that factor is estimated from inflation and contains the response being instrumented: the first stage falls to 8.8 and the set includes zero. Two caveats. At an inflation rank of two the first stage collapses, so the result rests on the paper’s rank choice; and the persistent level of u_t as instrument returns 0.08, below the innovation version, consistent with inflation responding to tightness with a lag, so the static IV is an impact response.

TABLE A.18: Instrumenting the synchronized component of PLS

	IV	First-stage F	Anderson–Rubin 95% set	OLS
Controls, no outcome factor	0.189	10.9	$[0.06, 0.36]$	0.105
Controls, outcome factor fixed	0.123	8.8	$[-0.08, 0.22]$	0.124
No controls, no outcome factor	0.371	14.3	$[0.23, 0.60]$	0.216

Notes: 2SLS regressions of PPI inflation on synchronized and idiosyncratic PLS components (idiosyncratic treated as exogenous), with sub-sector and country effects. Featured rows add the expectations proxy and baseline controls. Instrument: AR(1) innovation in the internal aggregate shock $\hat{u}_t$ interacted with pair exposure $\hat{\mu}_{ij}$; inflation rank one. First-stage F and Anderson–Rubin sets use date-clustered standard errors. For comparison, using the same innovation to instrument euro area-average tightness in the 88-quarter aggregate series yields first-stage $t = 1.6$ without controls and 0.8 with controls (Newey–West, 4 lags). The set is computed on a grid from -1 to 1.5 in steps of 0.01 . OLS uses the same specification without the instrument. 9,035 observations, 2003Q2–2024Q4.

D.7 Complements to the Robustness Checks

[↔ Back to reference in main text](#)

Robustness of the Component Slopes

The featured specification reports two component slopes, so its robustness must be assessed on both. Table A.19 re-estimates column (8) of Table 9 under the perturbations of this appendix – winsorizing, a post-pandemic indicator, lagged components and an alternative

supply-shortage control – re-running the full two-stage construction (factor extraction, loadings and sign normalization) inside every variant, alongside the conventional time fixed effects design of Table 8, column (4), on the same variant. Both component slopes are stable, and so is their ordering over the conventional estimate: α_x stays roughly twice the time fixed effects slope and α_c several times larger still, in every row.

TABLE A.19: Robustness of the Component Slopes

	Time FE	Common (α_c)	Idiosyncratic (α_x)	Obs.
Baseline (column (8))	0.016*** (0.006)	0.124*** (0.009)	0.043*** (0.007)	9,094
Winsorized 1/99	0.022*** (0.005)	0.122*** (0.009)	0.047*** (0.007)	9,094
Post-Covid dummy	0.016*** (0.006)	0.086*** (0.009)	0.033*** (0.006)	9,094
Lagged 1 quarter	0.019*** (0.006)	0.133*** (0.010)	0.041*** (0.008)	8,833
Lagged 2 quarters	0.021*** (0.006)	0.138*** (0.011)	0.038*** (0.009)	8,686
Capacity constraints	0.021*** (0.006)	0.123*** (0.009)	0.044*** (0.007)	9,093

Notes: Each row re-estimates the featured specification – column (8) of Table 9: interactive fixed effects at rank one with sub-sector and country effects, the baseline controls and the expectations proxy – on the indicated variant, with the two-stage decomposition re-estimated within the variant. The Time FE column reports the conventional design of Table 8, column (4), on the same variant. Winsorizing is at the 1% and 99% levels of inflation and PLS, applied before the decomposition. The lag rows lag both components (and, for the Time FE column, PLS) by the stated number of quarters. The post-Covid row adds a post-2021Q1 indicator, identified in the interactive specification, which carries no period dummies. The capacity row replaces reported supply shortages with capacity constraints, orthogonalized to PLS as in Appendix D.1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Common Estimation Sample and Weighted Granularity

Tables A.20 and A.21 report the weighted granularity diagnostics discussed in Section 7: Herfindahl indices and top- k cumulative shares under official EU Commission output weights, and the invariance of the estimated common factor of slack when the largest pairs by economic weight are removed and the two-stage construction is re-run on the reduced panel.

Dynamic Feedback: A Lead Test and a Reverse Regression

[↔ Back to reference in main text](#)

TABLE A.20: Granularity With Economic Weights

Equal-weighted HHI ($1/N$)	0.0081
Weighted HHI	0.0215
Largest pair share	0.0623
Top-1 / 5 / 10 / 24 cum. shares	0.06 / 0.24 / 0.37 / 0.59

Notes: Official EU Commission weights; pair weight = country weight $\times$ sector weight, renormalized over the common-sample pairs. The weighted HHI is the sufficient statistic of Assumption 1; per Gabaix (2011), a small N^{-1} does not by itself imply a small weighted HHI.

TABLE A.21: Leave-top- k by economic weight: the estimated common factor is unchanged

	(1)	(2)	(3)	(4)	(5)
	All	Drop top 1	Top 5	Top 10	Top 24
$ \text{corr}(\hat{F}_{\text{sub}}^x, \hat{F}_{\text{full}}^x) $	1.000	1.000	0.995	0.992	0.988
Observations	9094	9007	8656	8248	7131

Notes: Units ranked by official EU Commission output weights (country $\times$ sector); weighted HHI and top- k shares are reported in Table A.20. Each column drops the indicated number of largest pairs and re-runs the full two-stage construction – pair demeaning, factor extraction and loadings – on the reduced panel; the reported statistic is the absolute correlation, over common dates, between the re-estimated common factor of PLS and the full-sample factor.

Lagging PLS breaks contemporaneous feedback from inflation to reported shortages, but it does not by itself address dynamic feedback, anticipatory reporting – firms reporting shortages because inflation has been high – or a persistent omitted shock driving both series. This appendix reports two exercises that speak to those channels.

The first is a lead test: we add PLS led h quarters to the baseline IFE specification,

$$\pi_{ijt} = \alpha x_{ijt} + \beta_h x_{ij,t+h} + \lambda_{ij} F_t^y + \gamma_i + \tau_j + \delta \Gamma_{ijt} + \eta_{ijt}, \quad (27)$$

for $h = 1, 2, 4$. Conditioning on current PLS tests whether future shortages provide additional information about inflation beyond current tightness. Without this control, PLS persistence can generate an association between current shortages and past inflation even in the absence of reverse causality.

Panel A of Table A.22 reports $\hat{\beta}_h$. At $h = 4$ it is 0.005 and insignificant. This result provides limited evidence of a persistent omitted shock affecting both series. At $h = 1$ and $h = 2$ the lead coefficients are positive and significant, 0.035 and 0.024, against contemporaneous coefficients of 0.070 and 0.082.

TABLE A.22: Lagged-Inflation Placebo

	$h = 1$	$h = 2$	$h = 4$
<i>Panel A. Lead test. Outcome: π_{ijt}</i>			
PLS $_{t+h}$ (lead)	0.035 (0.006)	0.024 (0.005)	0.005 (0.004)
PLS $_t$	0.070 (0.006)	0.082 (0.006)	0.100 (0.006)
Observations	8833	8686	8451
<i>Panel B. Outcome: x_{ijt}; regressor: $\pi_{ij,t-h}$</i>			
Inflation $_{t-h}$	0.037 (0.013)	0.023 (0.012)	0.010 (0.013)
Observations	8833	8686	8451

Notes: Both panels use the baseline IFE specification of Section 7.2: one estimated factor with sub-sector $\times$ country loadings, additive sub-sector and country effects, and the baseline control set. Panel A adds PLS led h quarters to the baseline regression. The contemporaneous coefficient is reported alongside for comparison with the baseline of 0.093. Panel B reverses the regression and asks whether past inflation predicts current reported shortages, the anticipatory-reporting channel stated directly. Robust standard errors are reported in parentheses.

We do not read the short-horizon leads as evidence of contamination, and we do not claim they exclude it. A forward-looking Phillips curve predicts this sign: current inflation depends on expected future marginal cost, so tightness realized next quarter should help predict inflation this quarter, and the overlap between a quarterly survey and a quarterly price flow works in the same direction. The test cannot separate that structural channel from anticipatory reporting at short horizons. What it does establish is that whatever links the two series has died out within a year.

Panel B puts the anticipatory-reporting question the other way round: does past inflation predict currently reported shortages? The coefficients are 0.037, 0.023 and 0.010 at $h = 1, 2, 4$, with t -statistics of 2.9, 1.9 and 0.8 – small relative to the baseline slope and decaying to insignificance within a year. On this evidence reported shortages are not substantially a lagged echo of the inflation the specification sets out to explain.

Non-Linearity in the Phillips Curve

[↔ Back to reference in main text](#)

This appendix collects the implementation details and full results behind Section 7.5: the piecewise and state-dependent estimates on the two components of tightness, the random-forest estimation of the sub-sector-country thresholds, a simulated validation, robustness to

alternative thresholds and a block bootstrap that re-estimates the factor structure on every resample.

Specifications. Both specifications put the non-linearity on the components of tightness, and both are estimated on one sample, so the two columns are comparable. We do not report a piecewise fit in raw PLS: a kink in raw PLS is a kink in $\mu_{ij}F_t^x + \xi_{ijt}$ jointly, so it cannot say which component the steepening belongs to – exactly the conflation Lemma 2 warns about, which is the distinction examined in this appendix. Column (1) gives the idiosyncratic term its own kink, with thresholds trained on that component; the synchronized component enters linearly, because its structural reading is the conflated projection of Step 6 and a non-linearity in α_c estimated here would not be interpretable without an aggregate instrument. Column (2) interacts both components with an indicator for the pair being above its own threshold – the pair-level counterpart of the aggregate-state test of Section 7.5, and the specification that asks whether a pair’s own tightness changes its own slope, as opposed to whether the economy being tight changes it.

TABLE A.23: Non-Linearity in the Two Components of Tightness

	(1) Kink in ξ	(2) $\times$ tight state
Common (α_c)	0.157*** (0.008)	0.091*** (0.016)
Idiosyncratic (α_x)	0.023** (0.010)	0.056*** (0.011)
Kink in ξ	-0.019 (0.013)	
Common $\times$ tight (γ_c)		0.101*** (0.018)
Idiosyncratic $\times$ tight (γ_x)		-0.061*** (0.011)
Observations	8,979	8,979

Notes: Interactive fixed effects at rank one with sub-sector and country effects, the baseline controls and the expectations proxy. Column (1) gives the idiosyncratic component a kink at a threshold (Appendix D.7). Column (2) interacts both components with an indicator for the pair being above its threshold; the main effect of the indicator is included. The level coefficients in column (2) are normal-state slopes and are not directly comparable with the unconditional estimates of Table 9. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results. Neither specification indicates a steeper idiosyncratic response above the threshold. In column (2), the synchronized slope increases while the idiosyncratic slope declines. Defining tight states at the pair level therefore yields the same qualitative conclusion as the

aggregate-state analysis in Section 7.5: the estimated steepening is concentrated in synchronized tightness.

Inference. The thresholds are estimated rather than known, and the factor is estimated too, so the analytic standard errors above understate the true uncertainty. Table A.24 reports a block bootstrap of column (2) that resamples pairs with replacement and re-runs the factor extraction and the loadings inside every replication. Both interaction terms survive it: the ninety percent percentile interval for γ_c is $[0.050, 0.152]$ and for γ_x is $[-0.093, -0.035]$, so the steepening of the synchronized response and the flattening of the idiosyncratic one are each distinguishable from zero, and from each other. The two level coefficients are also bounded away from zero. What this interval does *not* price is the estimation of the thresholds themselves, which are held at their full-sample values inside the bootstrap; the factor-orthogonal exercise of Appendix D.7 speaks to that dimension instead.

TABLE A.24: State-Dependent Component Slopes: Block Bootstrap

	Estimate	Boot. SD	90% interval
Common (α_c)	0.091	0.027	[0.048, 0.136]
Idiosyncratic (α_x)	0.056	0.018	[0.030, 0.088]
Common $\times$ tight (γ_c)	0.101	0.031	[0.050, 0.152]
Idiosyncratic $\times$ tight (γ_x)	-0.061	0.018	[-0.093, -0.035]
Replications	200		

Notes: Column (2) of Table A.23. The bootstrap resamples sub-sector-country pairs with replacement, $B = 200$, all of which returned usable estimates, and re-estimates the factor and the pair loadings inside every replication. *Boot. SD* is the standard deviation across replications and the interval is the 90 percent percentile interval.

Random Forest Threshold Detection Procedure: Thresholds' Estimation. To identify the potential threshold in the relationship between labor shortages and inflation, we use a model-free, data-driven approach. Specifically, we estimate thresholds using a random forest regressor⁴⁶ trained separately on each sub-sector-country pair. Each forest consists of 500 decision trees, where every tree is fit to a bootstrap resample of the data. Within a single decision tree, the algorithm recursively searches for the best split points in the predictor space. When considering the relation between labor shortages and inflation, the algorithm tests candidate splits at which the relation may change and, at each step, the

⁴⁶Random forests are widely used in modern machine learning because they handle high-dimensional data, reduce overfitting and work well with minimal tuning – unlike e.g. neural networks, which are highly sensitive to hyperparameter choices.

algorithm selects the split that yields the greatest improvement in prediction accuracy⁴⁷. This process is repeated recursively for each branch of the tree.

Each split on PLS can be interpreted as a candidate threshold, which is the point at which the slope of the inflation-slack relationship may change. Because random forests consist of hundreds of such trees, each fit on a slightly different sample and, by allowing different branches, they generate a distribution of split points. From this distribution, we take the median PLS split as our estimate of the threshold for that sub-sector-country combination. Intuitively, this aggregates the “votes” of many trees about where the relationship between labor shortages and inflation changes most sharply.

The median split provides a data-driven threshold based on predictive fit. The simulation below illustrates its performance under a piecewise-linear relationship. We consequently treat x_{ij}^* as a generated regressor throughout, and support the non-linear results with two exercises: the bootstrap of Table A.24, which re-estimates the factor structure on every resample so the reported intervals carry the uncertainty of the decomposition itself; and the factor-orthogonal thresholds of Table A.25, which re-trains the thresholds on a series from which the common component has been removed.

This procedure has a natural economic interpretation within our framework. When PLS is below the threshold, the Phillips curve is relatively flat, indicating that inflation responds little to variation in labor market slack. Once PLS rises above the threshold, the Phillips curve steepens and inflation becomes highly sensitive to a tighter labor market.

The advantage of this method is that it does not impose any a priori functional form, enabling the data to inform the choice of the threshold, which we then use in the piecewise-linear specification. To the best of our knowledge, the use of random forests to estimate thresholds, at least in a Phillips curve setting, is new in the literature.

In the section below we show an example of this procedure on simulated data that share similar characteristics with our empirical sample. The economic content of the thresholds is evidenced by the estimated non-linearity itself and by its stability when the thresholds are re-estimated on factor-free PLS (Table A.25).

A Simple Illustration of the Random Forest Detection Procedure. This section gives a step-by-step explanation of the schematic example reported in the Figure A.6 below. The purpose of the illustration is pedagogical: it shows that the proposed data-driven

⁴⁷Equivalently, the largest reduction in mean squared error or increase in explained variance.

procedure recovers a meaningful threshold in a simple data generating process that mimics relevant features of the empirical panels. For this exercise, we generate a right-skewed distribution for the perceived labor shortages using a Gamma draw in order to have a distribution similar to the one observed in the empirical data. Moreover, to mimic the empirical feature that a non-negligible fraction of panels report zero or near-zero PLS, we force 10% of observations to be exactly zero. The true threshold is set to be equal to 10. Inflation is generated as the piecewise linear function of PLS

$$y = \begin{cases} 2.0 + 0.10 \cdot \text{PLS} + \varepsilon, & \text{if } \text{PLS} < 10 \\ 2.0 + 0.10 \cdot 10 + 0.20 \cdot (\text{PLS} - 10) + \varepsilon, & \text{if } \text{PLS} \geq 10 \end{cases}$$

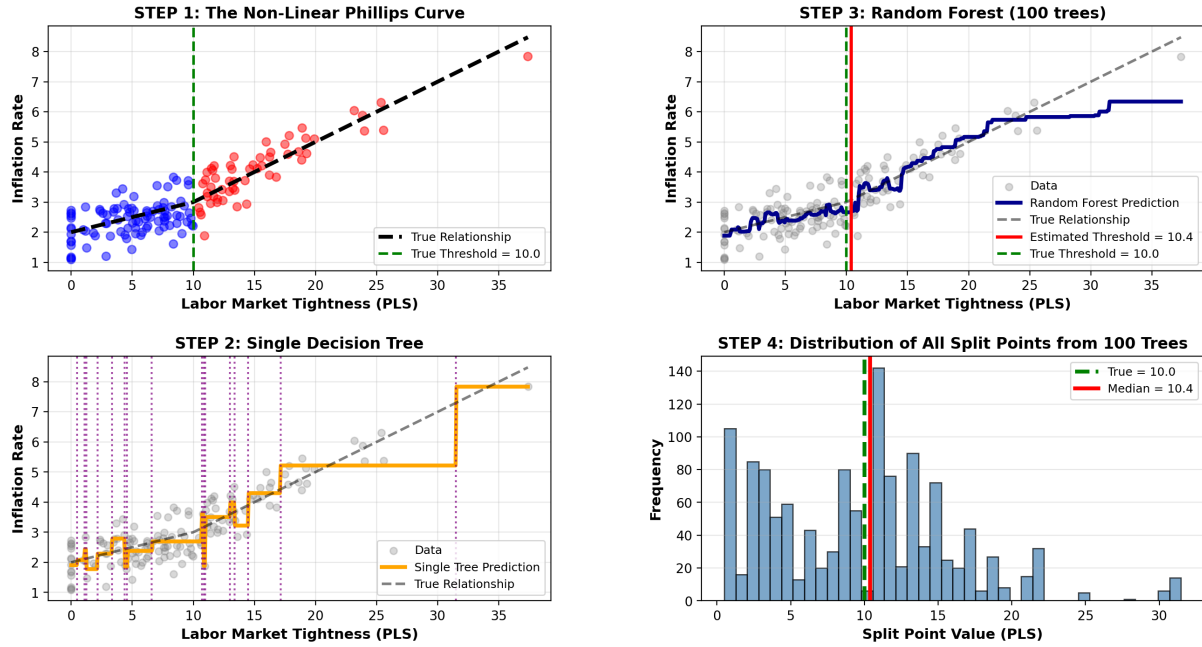
where $\varepsilon \sim N(0, 0.5^2)$. We fit a random forest similar to what we estimate in the data with each tree fitted to a bootstrap sample of the data and grows by recursive partitioning. A single decision tree searches over candidate splits of the form $\text{PLS} < \text{PLS}^*$ (for many candidate PLS^* values drawn from observed PLS) and selects the split that maximizes the reduction in predictive error. The chosen split is applied and the procedure repeats within each resulting node. Every internal node that splits on PLS provides a candidate split point (PLS^*) that marks a potential change in the slope of the inflation–PLS relationship.

After fitting the forest we iterate over every tree and every internal node. For nodes that split on the PLS feature we collect the corresponding threshold value. We will have a vector therefore that contains all split points on PLS found across all trees in the forest. Because each tree is grown on a different bootstrap sample and can split the predictor space differently, the vector contains a sample of plausible threshold candidates.

Finally, we take the median of that vector of candidate thresholds as the point estimate of the threshold for the example panel. The median is robust to outlying splits (for instance splits near zero that arise because many observations concentrate there) and corresponds to the majority “vote” of the ensemble about where the relationship changes.

The four panels in Figure A.6 visualize and schematize this process. In this illustrative example the forest yields a median split close to the true break (median close to 10.4 vs true 10) and the ensemble prediction reproduces the piecewise pattern. The single tree is noisier and its many candidate split points (purple lines) illustrate why an ensemble vote is preferable. The histogram in Panel 4 provides a direct, data-driven way to communicate both the point estimate (median). The exact same extraction and aggregation steps are applied to each empirical sub-sector-country panel in the dataset.

FIGURE A.6: Step-by-Step Random Forest Procedure on Simulated Data



Notes: The figure shows the procedure used to estimate thresholds for each sub-sector-country pair.

Factor-Orthogonal Thresholds

The thresholds are estimated on raw PLS, which contains the common component that the rest of the paper identifies as the confound. If the tightness episodes that determine above-threshold status are themselves driven by the aggregate, then the correlation between the common factor of PLS and the above-threshold share documented in Section 7.5, and with it the state dependence of Table A.23, could encode the factor rather than a structurally non-linear Phillips curve.

We re-estimate the thresholds using PLS net of its estimated synchronized component, $x_{ijt} - \hat{\mu}_{ij} \hat{F}_t^x$, maintaining pair-specific means. We then compute again column (2) of Table A.23 with these thresholds, continuing to define tight states using observed PLS. The revised thresholds reclassify approximately 10 percent of the observations.

On the same estimation sample, the results remain qualitatively unchanged (Table A.25): the synchronized slope increases in tight states, while the idiosyncratic slope declines. The estimated non-linearity is therefore robust to removing the common component from the series used to estimate the thresholds.

TABLE A.25: State Dependence under Factor-Orthogonal Thresholds

	Common (α_c)	Idiosyncratic (α_x)	Common $\times$ tight (γ_c)	Idiosyncratic $\times$ tight (γ_x)
Original thresholds	0.091*** (0.016)	0.056*** (0.011)	0.101*** (0.018)	-0.061*** (0.011)
Factor-orthogonal thresholds	0.070*** (0.015)	0.050*** (0.011)	0.134*** (0.018)	-0.054*** (0.010)
Observations	8,979	8,979	8,979	8,979

Notes: Both rows are column (2) of Table A.23 – interactive fixed effects at rank one with sub-sector and country effects, the baseline controls and the expectations proxy, with measured tightness entering as its two components and both interacted with an indicator for the pair being above its threshold. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

From Tight Labor Markets to Wages

The canonical view on slack and inflation, originating with Phillips (1958), holds that tight labor markets exert upward pressure on wages and, ultimately, on prices: when labor demand exceeds supply, firms offer higher wages to attract workers, and these cost increases are – at least partially – passed through into output prices (see, e.g., Ampudia et al., 2024).

TABLE A.26: Phillips Curve Estimates

	<i>Wages</i>
PLS	0.036** (0.015)
Constant	2.943*** (0.421)
Observations	827

Notes: Fixed effects (country level) and controls outlined in Section 7.2 are included. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Afrouzi et al. (2024) recently proposed a complementary mechanism where, under nominal wage rigidity, price level increases erode real wages and induce on-the-job search, raising job-to-job transitions and posted vacancies. Tightness indicators based on vacancies may therefore partly reflect the consequences of inflation rather than its causes. Our setting offers a useful, though partial, perspective on this point, since PLS relies on firms’ reported hiring difficulties rather than vacancy posting alone. As a simple check, we examine the pass-through from perceived shortages to wage growth. Table A.26 shows that increases in

PLS are associated with statistically significant wage increases⁴⁸. This is consistent with perceived shortages capturing economically meaningful labor market constraints, alongside any role for the search-based channel of Afrouzi et al. (2024). A sharper decomposition of the two mechanisms is a promising avenue for future research.

D.8 Im, Pesaran and Shin (2002) Unit Root Test

TABLE A.27: Unit Root Test Results

Variable	Z-stat	p-value
Inflation	-13.418	0.000
PLS	-34.163	0.000
Supply Shortages	-48.012	0.000
Insufficient Demand	-37.850	0.000
Other Constraints	-54.299	0.000
Financial Constraints	-47.626	0.000
WEO Inflation Expectations	-5.315	0.000
Energy Prices	-27.042	0.000
NEER	-99.050	0.000

Notes: In the Im et al. (2003) unit root test the null hypothesis is that all panels contain unit roots, while the alternative hypothesis is that at least some panels are stationary.

D.9 Alternative Standard Errors

Alternative standard errors. Table A.28 compares covariance estimators applied to the same fitted specifications and estimation sample. Relative to panel clustering, two-way clustering increases standard errors for the synchronized slope by approximately 40 percent, while those for the idiosyncratic slope change little. Both slopes remain statistically significant throughout.

Panel-clustered standard errors are also close to those from the bootstrap that re-estimates both stages (Table A.29). Their similarity suggests that within-pair dependence accounts for much of the increase relative to heteroskedasticity-robust standard errors.

⁴⁸The analysis is conducted at the country level, as wage indices are not available at the NACE-2 level.

TABLE A.28: Point Estimates Are Invariant to the Covariance Estimator

	(5) IFE	(6) FE	(7) Sec. × Cty	(8) Infl. Exp.
Common (α_c)	0.163***	0.151***	0.147***	0.123***
Robust (HC) SE	(0.009)	(0.009)	(0.009)	(0.009)
Cluster: panel	(0.020)	(0.021)	(0.020)	(0.019)
Cluster: panel and time	(0.028)	(0.029)	(0.029)	(0.027)
Idiosyncratic (α_x)	0.039***	0.059***	0.053**	0.044**
Robust (HC) SE	(0.006)	(0.009)	(0.010)	(0.008)
Cluster: panel	(0.014)	(0.021)	(0.021)	(0.017)
Cluster: panel and time	(0.015)	(0.022)	(0.022)	(0.018)
Observations	9094	9094	9094	9094

Notes: Each column reports a single fit of the stage-wise decomposition – the specifications of Table 9, columns (5)–(8) – with the three covariance estimators applied to that same fit. Coefficients and standard errors are computed from the concentrated regression at the converged factor, written out with explicit loading dummies, which reproduces the interactive-effects coefficients up to the estimator’s own convergence slack; they can accordingly differ from Table 9 in the third decimal. Stars refer to the most conservative of the three, the two-way clustered standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.29: Phillips Curve Estimates: Panel Clustered SE

	Factor Loadings				
	(5) IFE	(6) FE	(7) Same	(8) Sec. × Cty	(9) Infl. Exp.
Common (α_c)	0.163*** (0.020)	0.153*** (0.021)	0.059*** (0.020)	0.149*** (0.020)	0.124*** (0.019)
Idiosyncratic (α_x)	0.039*** (0.014)	0.058*** (0.021)	0.034** (0.015)	0.052** (0.021)	0.043** (0.017)
Observations	9094	9094	9094	9094	9094

Notes: Specifications from Table 9, with standard errors clustered by sub-sector–country pair in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix to “Labor Market Shortages and the Phillips Curve: Identification under Heterogeneous Exposure to Common Shocks”

Cristian Arcidiacono

OA.1	Online Appendix: Complements to the Method	117
OA.1.1	Detailed Links with Previous Literature	117
OA.1.2	Extension: Diff-in-Diff with Heterogeneous Factor Exposure	122
OA.1.3	Rank Estimation (Number of Factors)	127
OA.2	Online Appendix: Complements to the Application	131
OA.2.1	Time Series of PLS by Country and Sub-Sector	131

OA.1 Online Appendix: Complements to the Method

OA.1.1 Detailed Links with Previous Literature

The problem this paper addresses – contamination of a panel regression by a common factor with heterogeneous unit loadings – is a special case of an identification problem studied in the recent literature: how to recover a micro-local elasticity when units respond heterogeneously to aggregate, general-equilibrium objects, so that the time fixed effect no longer absorbs all equilibrium feedback. The standard two-way design – canonical since [Nakamura and Steinsson \(2014\)](#) and codified by [Chodorow-Reich \(2020\)](#) – is subject to the identification problems formalized in Section 3, and the responses in the literature can be grouped by the external ingredient each introduces to restore identification. IFE requires no such ingredient: where the other approaches import an observed shock, observed exposure weights or a fully specified equilibrium model, it identifies the slope internally, from the granularity of the panel.

Balancing and design-based approaches. One strand identifies by comparing units appropriately *balanced* in their exposure rather than by relying on exogenous time-series shocks ([Borusyak et al., 2022](#); [Majerovitz and Sastry, 2023](#)): the shift-share theory of [Goldsmith-Pinkham et al. \(2020\)](#), the reweighting procedure of [Arkhangelsky and Korovkin \(2024\)](#), and the targeting-rule estimator of [Manuel \(2026\)](#), which extends the logic to dynamics and multiple general-equilibrium objects. All rely on observed unit characteristics or shares that span the loading heterogeneity: collecting the observed exposures in a matrix S with projection P_S , the balancing route recenters the slope under a weaker condition than spanning – the residual exposure $(I - P_S)\mu$ need only be mean-independent of the outcome loading ([Manuel, 2026](#), Proposition 5) – but it removes the variance floor of Lemma 3 only if $(I - P_S)\lambda = 0$, and leaves a floor exactly to the extent of the unspanned residual; this is the “recentered only” entry of Table OA.1. IFE dominates where that condition is hardest to meet – loading heterogeneity reflecting unobservables such as technology or market structure – because it estimates the loadings rather than proxying for them; the ordering reverses when the panel is not granular, where a balancing scheme built on credibly exogenous shares is the more reliable tool. IFE trades a data requirement, observed exposure, for a structural one, granularity.

External-instrument and time-series approaches. A second strand imports an external, identified shock to the aggregate and nets out its general-equilibrium response. [Don-](#)

aldson (2025) shows that the control-function correction commonly used to address time fixed effect bias generally fails in dynamic settings when units respond heterogeneously to an aggregate that moves in general equilibrium, and proposes a decomposition that recovers the micro-local term from time-series evidence on identified shocks, in the spirit of McKay and Wolf (2023); Wolf (2023) provides the demand-equivalence conditions under which such evidence recovers the general-equilibrium intercept missing from cross-sectional designs⁴⁹. Identification runs through the time-series dimension, conditional on a clean external shifter, and the route is built for the regime in which the aggregate’s response to individual units is first-order – exactly where IFE’s granularity assumption fails. The two are therefore close to mutually exclusive in their domains: IFE applies when no credible external shock exists and feedback is diffuse across many small units; the external-instrument route when a clean shifter is in hand and no unit is negligible. The informational requirements of that route are spelled out by Manuel (2026, Appendix C.1): the complete list of general-equilibrium aggregates through which the shock transmits, and enough valid instruments to hold every one of them fixed at every horizon.

Model-based and inference-focused approaches. A third group writes down a fully specified general-equilibrium model and estimates a model-consistent reduced form (Adão et al. 2019), or translates cross-sectional estimates into aggregate effects through factor structures or SVAR-based schemes (Sarto 2025; Matthes et al. 2025): the strongest identification available, at the cost of exclusion restrictions on a policy block or of validity holding only within the assumed model class. In the classification of Manuel (2026), who separates the recovery of direct effects from the recovery of the missing intercept and calls the two complementary, the micro-local slope of this paper belongs to the first task and the synchronized slope of Section 7.4 to the second, both from one panel. IFE occupies the opposite corner, imposing almost no structure beyond a factor representation, though it is correspondingly silent on mechanisms the factor structure does not encode. Canova (2024) documents a complementary failure of cross-regional designs – heterogeneity in the *dynamics* of local responses rather than in loadings. Distinct again is the inference-oriented work of Almuzara and Sancibrián (2024) and the exposure-robust standard errors of Adão et al. (2019), which addresses related inference problems in shift-share designs. These methods concern inference conditional on identification: under heterogeneous loadings the time-dummy estimand

⁴⁹Donaldson (2025) reports a simulation in which IFE fails to recover the causal coefficient. This is a boundary case of Assumption 4 rather than evidence against factor-based identification: her regressor is a pure exposure–shock product with no variation outside the outcome’s factor space, so $\inf_F D(F) = 0$ and, as she notes, no increase in rank can manufacture what is missing. This panel supplies that variation twice over – idiosyncratic slack variation, and a common component only weakly aligned with the inflation factor.

is itself the wrong object (Appendix B.4), so correct coverage of the wrong estimand is not a resolution, though the two exercises are complementary once the estimand is fixed.

These approaches differ in their identifying information and assumptions. Designs based on observed exposures or external shocks depend on the validity of those measures. IFE instead estimates latent exposures from the panel under a factor structure and exogeneity conditions. Granularity limits feedback from individual units to the aggregate, supporting identification of the micro-local slope without an external instrument. The approaches therefore rely on distinct assumptions. Table OA.1 summarizes these differences, while Table OA.2 outlines the settings in which each approach is more appropriate. Section 7 assesses concentration in the euro area panel.

Whichever route is taken, standard errors should be clustered by unit. Under period dummies that is the only estimator that prices the floor of Lemma 3, and the interval it returns is floor-bound. Under the observable-control routes the floor is removed in proportion to the exposure the controls capture – what the outcome’s factor structure leaves in the residual is then within-unit dependence, ordinary and priceable, with a floor surviving only through the exposure the controls miss – and unit clustering prices that dependence. In neither case does a time-pooled correction help. The alternative is the design-based view, in which the aggregate shocks are the source of randomness and clustering runs by period (Almuzara and Sancibrián, 2024), the choice Manuel (2026) adopts in his fiscal application; under heterogeneous exposure left in the residual, Lemma 3 says the two views price different objects.

Granular instrumental variables. The granularity condition also connects this paper to the granular-origins literature, read in reverse. Gabaix (2011) showed that concentration lets idiosyncratic shocks drive aggregate fluctuations, and Gabaix and Koijen (2024) turn that into an estimator: where a few large units dominate, their idiosyncratic shocks instrument for the aggregate, with the excess Herfindahl governing strength. Our assumption reads the same statistic in the opposite direction – it requires the *absence* of concentration, so that idiosyncratic variation is asymptotically orthogonal to the aggregate and identifies the micro-local slope from within the panel. Alvarez-Blaser et al. (2025) provide empirical evidence for both ends of this axis: with barcode-level prices for sixteen countries they model disaggregated inflation exactly as equation (1) without the regressor, documenting that the loading heterogeneity our analysis treats as the confound is a robust feature of disaggregated price data, dominant at the product-category level – the closest analogue to our sector×country panel. Their exercise is an accounting decomposition of inflation itself;

TABLE OA.1: Identification Routes for the micro-local Slope

	IFE (this paper)	Balancing / design-based	External instrument
Identifying variation	Within-panel	Cross-sectional	Time-series
Factor structure	Estimated	Controlled for (requires $\rho_{\lambda\mu} = 0$ given controls)	Absorbed via shock structure
Unit chars. needed?	No	Yes (μ_i)	Yes (weights)
Identified shock needed?	No	No (unpredictable shifter)	Yes
Granularity required?	Yes	Assumed (continuum)	No
GE feedback handled?	Yes	Yes, given the targeting rule	Yes
Inference	Repaired	Recentered only	Recentered only

Notes: All routes target the micro-local slope α in $y_{it} = \alpha x_{it} + \lambda_i F_t + \gamma_i + \varepsilon_{it}$, $x_{it} = \mu_i F_t + \xi_{it}$. The two comparison columns summarize classes of methods rather than single papers; representative instances are Goldsmith-Pinkham et al. (2020), Arkhangelsky and Korovkin (2024) and Manuel (2026) for the balancing route; Donaldson (2025) and McKay and Wolf (2023) for the external-instrument route. “Controlled for” means the method removes the $\lambda_i F_t$ confound through observed characteristics and requires the residual exposure to be mean-independent of the outcome loading given those controls (Manuel, 2026, Proposition 5); the balancing route needs no identified shock, only a shifter unpredictable from its own past (Manuel, 2026, Proposition 6), and its baseline continuum of units builds granularity in. “Granularity” (Assumption C5) means each unit’s idiosyncratic variation contributes $O(N^{-1})$ to the aggregate.

the consequences of the loading structure for slope estimation – our subject – are left open. Chaudhary et al. (2026) have since relaxed the concentration requirement: under precision- and size-weighted moment conditions, their optimal GIV draws its strength from the pooled idiosyncratic variance rather than from the excess Herfindahl, so the estimator does not need a granular panel. The relaxation does not reach our object, however. The GIV uses idiosyncratic shocks to instrument the aggregate and so identifies the units’ response to it – the general-equilibrium feedback that Assumption 1 bounds – not the slope on the common shock; and when the common factor must be estimated from within the panel, identification is lost once the units’ loadings on the aggregate lie in the span of the factor loadings, a direct consequence of their Proposition A.2. In equation (2) the natural feedback channel runs through the same exposure μ_i that carries the common shock – a unit’s tightness responds to aggregate tightness in proportion to its exposure to it – which is exactly that case.

TABLE OA.2: Choosing among designs for panels with common shocks

Route	Substitutes for the loadings	Appropriate when	Breaks when
Time dummies	Nothing; imposes homogeneous exposure	The response to common and to idiosyncratic movement coincide, $\alpha_c = \alpha_x$, so deleting the common part costs precision alone (Remark 1)	Loadings are dispersed: the estimate is biased whenever they covary with the outcome's (Lemma 1), and conventional standard errors understate regardless; clustering by unit prices the floor but cannot remove it (Lemma 3, Appendix B.4)
Interactive fixed effects (<i>this paper</i>)	Granularity — a property of the sampling frame, not of the researcher's information	Exposure is latent or incompletely observed and no unit is individually large (Assumption 1)	A few units carry a non-negligible share of the aggregate, or the regressor is noise-dominated (Experiment G6)
Balancing shift-share	/ Observed exposure shares	The shares are accurate and complete and the panel is <i>coarse</i> — the standard shift-share setting, where this route dominates outright (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022), as does synthetic control in static settings (Arkhangelsky and Korovkin, 2024)	The shares miss part of the exposure, $(I - P_S)\lambda \neq 0$
External instrument / GE decomposition	An observed aggregate shock	The identifying variation <i>is</i> an aggregate shock times exposure, so no increase in rank can manufacture it (Donaldson, 2025; McKay and Wolf, 2023)	No credible external shifter is available
Targeting-rule controls	Observable characteristics satisfying a targeting rule	Such characteristics exist and the rule holds (Manuel, 2026)	The rule fails, and the controls no longer span the confound

Notes: The routes are a substitution of assumptions rather than a ranking, and they do not nest. Each is diagnosable before estimation: granularity by the weighted concentration of the panel ($N \cdot \text{HHI}$, Assumption 1); share completeness by the span condition $(I - P_S)\lambda = 0$ stated above; the homogeneity that would justify time dummies by estimating α_c and α_x separately (Section 7.4). The last is the diagnostic this paper adds: where the two component slopes differ, time dummies change the estimand and not merely its precision, whichever alternative is then chosen.

OA.1.2 Extension: Diff-in-Diff with Heterogeneous Factor Exposure

This section shows that the three failures of Section 3 are not specific to the continuous-regressor panel: each has an exact counterpart in the workhorse difference-in-differences (DiD) design. We make no claim to a new DiD estimator. That heterogeneous exposure to latent common shocks can violate parallel trends is itself familiar – it is the factor-model motivation for synthetic control (Abadie et al., 2010) and for interactive-effects approaches to program evaluation (Gobillon and Magnac, 2016; Xu, 2017; Arkhangelsky et al., 2021; Callaway and Karami, 2023) – and the associated inference problem is the aggregate-shock/serial-correlation critique of Bertrand et al. (2004) and Arkhangelsky and Korovkin (2024). Manuel (2026) exhibits the general-equilibrium counterpart: parallel trends and no anticipation hold exactly, yet the difference-in-differences coefficient misses the direct effect because the treatment itself moves the aggregates to which units respond heterogeneously. The estimand failure needs no importing at all: its binary incarnation is the celebrated weighting problem of two-way fixed effects under heterogeneous treatment effects⁵⁰. The value of this section is unification: the bias and inference failures are exact special cases of Lemmas 1 and 3, recovered by degenerating the regressor to a treatment indicator, so the mechanism developed in the main text is design-independent rather than an artifact of the continuous-regressor setting.

Setup. Retain the outcome equation (1) but replace the continuous regressor with a treatment indicator. In the canonical two-group, two-period-block design,

$$y_{it} = \alpha D_{it} + \lambda_i F_t + \gamma_i + \varepsilon_{it}, \quad D_{it} = T_i P_t, \quad (28)$$

where $T_i \in \{0, 1\}$ marks treated units and $P_t = \mathbb{1}\{t > t^*\}$ marks post-treatment periods. Comparing with the framework of Section 3, the regressor equation (2) has not disappeared – it has degenerated: the regressor’s “loading” on the aggregate is now the group indicator T_i itself, and it is *observed*, because it is the design. What remains latent is exactly what was latent before: the outcome loading λ_i . DiD thus eliminates one of the two unobserved exposure objects of the main text and concentrates the entire identification problem in the other.

Proposition 2 (DiD as a Special Case of Lemma 1). *Consider the DGP (28) in a balanced*

⁵⁰See, among others, Goodman-Bacon (2021); de Chaisemartin and D’Haultfoeulle (2020); Sun and Abraham (2021); Callaway and Sant’Anna (2021).

panel. Let the loadings λ_i be i.i.d. across units with finite variance, with distribution allowed to depend on treatment status; let ε_{it} be mean-zero, independent across units and of (T_i, λ_i) and $\{F_t\}$; and let $\Pr(T_i = 1) \in (0, 1)$. Condition on the realized factor path $\{F_t\}_{t=1}^T$. Then, as $N \rightarrow \infty$ with T fixed, the two-way fixed effects estimator of α (equivalently, the DiD estimator) satisfies

$$\text{plim}_{N \rightarrow \infty} (\hat{\alpha}_{\text{DiD}} - \alpha) = \Delta_\lambda \cdot (\bar{F}_{\text{post}} - \bar{F}_{\text{pre}}), \quad \Delta_\lambda \equiv \mathbb{E}[\lambda_i \mid T_i = 1] - \mathbb{E}[\lambda_i \mid T_i = 0], \quad (29)$$

where $\bar{F}_{\text{post}}$ and $\bar{F}_{\text{pre}}$ are the factor means over the post- and pre-treatment windows. The collapsed pre/post DiD coefficient is therefore unbiased if and only if $\Delta_\lambda \cdot (\bar{F}_{\text{post}} - \bar{F}_{\text{pre}}) = 0$.

This block-averaged condition should be distinguished from parallel trends understood pathwise. The proof below equates $\hat{\alpha}_{\text{DiD}} - \alpha$ with a contrast of window averages, so it characterizes only the moment the two-period estimator uses. The date-by-date requirement is strictly stronger: for every relevant pair of dates (t, s) , either $\Delta_\lambda = 0$ or $F_t - F_s = 0$. A factor path that moves within the pre-period and within the post-period but has equal window means satisfies the displayed condition while violating parallel trends at every date. Since the discussion that follows reads pre-trends date by date, and the event-study illustration turns on exactly that reading, we use the pathwise notion there and the block-averaged one here.

Proof. In the balanced design the two-way demeaned treatment indicator factorizes exactly, $\tilde{D}_{it} = (T_i - \bar{T})(P_t - \bar{P})$, and by Frisch–Waugh $\hat{\alpha}_{\text{DiD}} - \alpha = \sum_{it} \tilde{D}_{it}(\lambda_i F_t + \varepsilon_{it}) / \sum_{it} \tilde{D}_{it}^2$. The numerator separates, $\sum_{it} \tilde{D}_{it} \lambda_i F_t = [\sum_i (T_i - \bar{T}) \lambda_i] [\sum_t (P_t - \bar{P}) F_t]$, and dividing numerator and denominator by NT gives $\widehat{\text{Cov}}(T_i, \lambda_i) \widehat{\text{Cov}}_t(P_t, F_t) / \widehat{\text{Var}}(T_i) \widehat{\text{Var}}_t(P_t)$ plus an ε -term that vanishes in probability. As $N \rightarrow \infty$, $\widehat{\text{Cov}}(T_i, \lambda_i) / \widehat{\text{Var}}(T_i) \rightarrow \Delta_\lambda$ (the regression of λ_i on a binary T_i recovers the mean gap), while the time-series ratio equals $\bar{F}_{\text{post}} - \bar{F}_{\text{pre}}$ identically. The parallel-trends claim follows from $\mathbb{E}[y_{it}(0) - y_{is}(0) \mid T_i = 1] - \mathbb{E}[y_{it}(0) - y_{is}(0) \mid T_i = 0] = \Delta_\lambda(F_t - F_s)$. $\square$

Equation (29) is Lemma 1's formula with every term replaced by its design counterpart: the latent loading covariance $\text{Cov}(\tilde{\mu}_i, \tilde{\lambda}_i)$ becomes the covariance between treatment status and the outcome loading, $\text{Cov}(T_i, \lambda_i) = \text{Var}(T_i) \Delta_\lambda$; the factor variance $\text{Var}(\tilde{F}_t)$ becomes the factor's pre/post drift; and $\text{Var}(\tilde{x}_{it})$ becomes the design variance $\text{Var}(T_i) \text{Var}(P_t)$. The economics of $\Delta_\lambda \neq 0$ is selection on exposure rather than selection on levels: unit fixed effects absorb permanently different units, but nothing in the design precludes a program being rolled out precisely where sensitivity to the aggregate cycle is highest – regions targeted because a downturn hits them hardest, banks regulated because their balance sheets amplify

the common financial factor. In such settings $\Delta_\lambda \neq 0$ by construction of the policy.

Pre-trend tests have no power against a quiescent factor. The loading gap Δ_λ produces divergent pre-trends only through pre-period movement in F_t : the group-level gap at dates $s < t \leq t^*$ is $\Delta_\lambda(F_t - F_s)$. If the factor is flat before t^* and moves afterwards – a crisis, a policy cycle turning, an energy shock – pre-trends are exactly parallel in expectation while identification fails by (29). This is distinct from the familiar low-power critique of pre-trend testing (Roth, 2022): the issue is not that the test is underpowered against a violation already present in the pre-period, but that the violation resides entirely in the post-period realization of the factor, where no pre-period test can see it.

The inference failure. The inference failure of Section 3.3 also carries over. Set $\Delta_\lambda = 0$, so Proposition 2 delivers a consistent point estimate. The residual still contains $(\lambda_i - \bar{\lambda})\tilde{F}_t$: a serially correlated component common to all units up to the loading. Its consequences in DiD are well documented under other names. Heteroskedasticity-robust standard errors that ignore the within-unit serial correlation of the factor component undercover severely – the critique of Bertrand et al. (2004) – and the simulation below, which follows the main text’s convention of heteroskedasticity-robust inference, finds coverage of 0.66 at $\Delta_\lambda = 0$. Clustering by unit repairs this only when the number of units informative about the loading gap is large; in the typical application with few treated units, the realized gap $\bar{\lambda}_1 - \bar{\lambda}_0$ is dominated by a handful of draws interacted with a single realization of the factor path, which is the inference problem that motivates Arkhangelsky and Korovkin (2024) and the synthetic-control literature. Viewed from the superpopulation perspective of Appendix B.4, the term $(\bar{\lambda}_1 - \bar{\lambda}_0)(\bar{F}_{\text{post}} - \bar{F}_{\text{pre}})$ is the DiD incarnation of the random conditional bias of Lemma 3.

Staggered adoption and the modern DiD corrections. With staggered timing, $D_{it} = \mathbb{1}\{t \geq G_i\}$ for cohort dates G_i , the relevant object becomes the covariance between adoption timing and the loading, and the same logic applies cohort by cohort. Two remarks are worth making. First, cohort-specific linear trends – the standard patch – do not resolve the problem, because a linear trend does not span a factor path: the confound is $\lambda_i F_t$, not $\lambda_i t$. Second, the taxonomy of the main text applies verbatim. The negative-weighting problem of heterogeneous treatment effects addressed by Callaway and Sant’Anna (2021) and Sun and Abraham (2021) is the binary incarnation of Failure II: those corrections repair *what is estimated* – the aggregation of heterogeneous α ’s – while maintaining parallel trends. Proposition 2 is Failure I: parallel trends fails itself, with α homogeneous. A design can suffer either alone or both at once, and the remedies are disjoint – which yields the practical

statement this section exists to make: the modern staggered-DiD estimators repair the estimand failure and neither of the other two. A design can pass every heterogeneity-robust check and still carry the factor-exposure bias of Proposition 2 and the variance floor of Lemma 3. A further parallel completes the mapping: the adoption share is the treatment’s own aggregate – the analogue of F_t^x – and treatment effects that depend on how many units are treated, through spillovers or general-equilibrium adjustment, are the binary form of the two-estimand structure of Remark 2: the own-treatment effect is identified from cross-sectional comparisons, the response to the adoption wave only from aggregate variation (Wolf, 2023).

Remedies. The IFE logic of the main text extends directly⁵¹. The assumptions the main text makes explicit have direct analogues: a long pre-treatment window plays the role of $T \rightarrow \infty$ in pinning down loadings (with short panels requiring the additional structure of Callaway and Karami, 2023), and Assumption 1 has a sharp counterpart in the size of the treated group. The counterpart should be stated carefully. With a single treated unit, *pooled regression-style* IFE cannot separate that unit’s loading from its treatment effect, and few-treated inference is compromised for the usual large-treated-group reasons. That is not the same as saying factor-based counterfactual estimation is unavailable: given a long pre-treatment window and an untreated donor pool that pins down F_t , the treated unit’s loading can be learned from its pre-treatment outcomes and post-treatment deviations then identify the effect – which is exactly what synthetic control and its factor-model relatives do. The limitation is specific to the pooled-regression implementation, and it is in that regime that synthetic-control weighting, rather than IFE regression, is the appropriate tool.

OA.1.2.1 Monte Carlo: Staggered DiD with Factor Exposure

DGP. We simulate a staggered design with four cohorts, calibrated to the conventions of Section 5: treatment dates at the quartiles $G \in \{25, 50, 75, \infty\}$, $N = 240$ (60 units per cohort), $T = 100$, $\rho_F = 0.80$, loading mean 0.5 and standard deviation 1. Early-treated cohorts have higher factor loadings; the parameter δ_λ governs the correlation between treatment timing and λ_i , so δ_λ plays the role of Δ_λ in Proposition 2. The true treatment effect $\alpha = 1$ is homogeneous, switching off the heterogeneous-effects channel by construction. Estimators: TWFE with heteroskedasticity-robust standard errors, IFE at rank 1, and TWFE with cohort-specific linear trends. Because the per-replication conditional bias $\Delta_\lambda(\bar{F}_{\text{post}} - \bar{F}_{\text{pre}})$ has mean zero across draws of a stationary factor path, the expected signature is not a

⁵¹See Gobillon and Magnac (2016); Xu (2017); Arkhangelsky et al. (2021); Abadie et al. (2010) for more details.

signed location shift in the TWFE mean but inflated dispersion and collapsing coverage – the random-conditional-bias pattern of Lemma 3.

Results. Table OA.3 ranges $\delta_\lambda \in \{0, 0.5, 1.0, 1.5\}$, and the results match the predicted signature exactly. The TWFE mean stays within 0.06 of the true $\alpha = 1$ at every δ_λ – the conditional bias averages out across factor draws – but its dispersion explodes, from a standard deviation of 0.11 at $\delta_\lambda = 0$ to 1.58 at $\delta_\lambda = 1.5$, while the mean reported robust standard error rises only from 0.04 to 0.08: an understatement factor of roughly twenty at the top of the range. Coverage falls accordingly, from 0.66 with *no* loading gap at all (the pure inference failure) to 0.09 at $\delta_\lambda \geq 1$. Cohort-specific linear trends do not repair the problem – a trend cannot absorb a factor path – with coverage falling from 0.71 to 0.08 across the range. IFE at rank 1 is centered on α throughout (absolute bias at most 0.002), its reported standard error matches the true sampling dispersion, and coverage sits at 0.94–0.95 in every cell: it absorbs $\lambda_i F_t$ whether or not λ_i correlates with treatment timing.

TABLE OA.3: Staggered DiD with Heterogeneous Factor Exposure

Estimator	Timing–loading correlation δ_λ			
	$\delta_\lambda = 0$	$\delta_\lambda = 0.5$	$\delta_\lambda = 1.0$	$\delta_\lambda = 1.5$
TWFE	0.999 {0.66}	1.043 {0.13}	0.981 {0.09}	1.055 {0.09}
IFE (rank 1)	1.000 {0.94}	1.000 {0.95}	0.998 {0.95}	1.001 {0.94}
TWFE + cohort trends	1.005 {0.71}	1.022 {0.18}	0.953 {0.11}	1.057 {0.08}
<i>Condition: staggered adoption, α homogeneous. {Coverage}</i>				

Notes: 500 replications. $\alpha = 1$, $N = 240$ (four cohorts of 60), $T = 100$, $\rho_F = 0.80$, treatment dates $G \in \{25, 50, 75, \infty\}$. Loadings $\lambda_i = 0.5 + \delta_\lambda s_g + \eta_i$ with cohort scores $s_g \in \{1.5, 0.5, -0.5, -1.5\}$ (earliest treated loads highest) and $\eta_i \sim \mathcal{N}(0, 1)$, so δ_λ indexes the treatment-timing–loading correlation of Proposition 2. TWFE standard errors are heteroskedasticity-robust, following the main text’s convention. The per-replication conditional bias $\Delta_\lambda(\bar{F}_{\text{post}} - \bar{F}_{\text{pre}})$ has mean zero across factor-path draws, so TWFE means remain near α while dispersion inflates and coverage collapses – the random-conditional-bias signature of Lemma 3. Cohort-specific linear trends do not span a factor path.

An event-study illustration: clean pre-trends, failed identification. Figure OA.1 turns the zero-signal case discussed after Proposition 2 into a picture. Two groups (120 treated at $t^* = 50$, 120 never treated) differ in their loadings by $\Delta_\lambda = 1$; the factor is *quiescent* before t^* ($F_t = 0$) and follows the calibrated AR(1) from t^* onward – an aggregate

shock arriving with the treatment. Because the pre-period factor does not move, every lead coefficient in the TWFE event study is centered exactly at zero in population, and the Monte Carlo means of the leads sit at zero to simulation error. A pre-trend test run on the design therefore has no systematic tendency to reject, however much data is available. Individual replications are a different matter: finite-sample leads still fluctuate with the idiosyncratic shocks, so a given replication can fail a conventional test at its nominal size. The point is not that every draw passes, but that the design carries no signal a pre-trend test could detect. The post-period coefficients are a different matter. Each example replication drifts persistently away from the true effect – the loading gap interacted with that replication’s factor realization – producing paths that to a researcher look like dynamic treatment effects building, decaying or reversing over time. The magnitudes are stark: the mean lead coefficient never exceeds 0.010 in absolute value, while the 90% Monte Carlo band at $k = +10$ is more than six units wide against a true effect of one. Averaged across replications the paths are centered near the truth (the conditional bias has mean zero over factor draws), but any *single* realized study can land far from it, with clean pre-trends offering no warning. The figure is the visual counterpart of the remark above: the parallel-trends violation lives entirely in the post-period realization of the factor, where no pre-period diagnostic can detect it.

OA.1.3 Rank Estimation (Number of Factors)

Consider a model similar to the one estimated for the NKPC application⁵²:

$$\pi_{ijt} = \alpha x_{ijt} + \lambda_{ij} F_t^y + w_{ijt}.$$

Bai and Ng (2002) estimate the number of factors by penalising a loss function with a term that increases in both panel dimensions, so that adding a factor is worthwhile only when it reduces the loss by more than the penalty charges. They propose two variants, the panel criterion (PC) and the panel information criterion (IC), which differ in whether the penalty is applied to the loss or to its logarithm:

$$\begin{aligned} \text{PC}(k) &= V(k, \hat{F}^k) + k \hat{\sigma}^2 g(N, T) \\ \text{IC}(k) &= \ln V(k, \hat{F}^k) + k g(N, T). \end{aligned}$$

⁵²This section largely comes from Ditzen and Reese (2023).

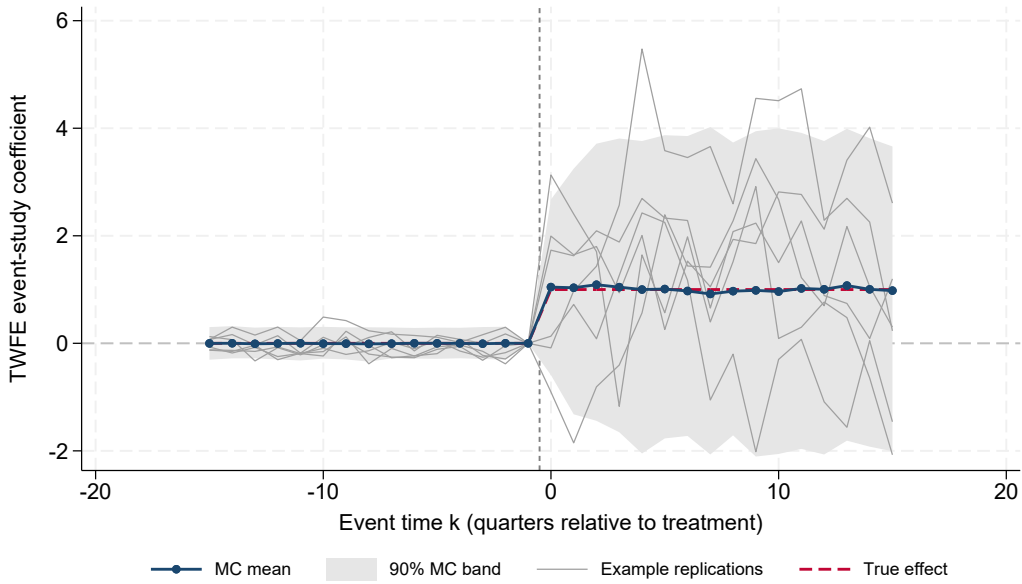


FIGURE OA.1: Event-Study Coefficients under a Quiescent-then-Active Factor

Notes: TWFE event-study coefficients ($k = -1$ omitted) from the DiD DGP with loading gap $\Delta_\lambda = 1$ and factor $F_t = 0$ for $t < t^*$, AR(1) with $\rho_F = 0.80$ thereafter. $N = 240$ (120 treated at $t^* = 50$), $T = 100$, window $k \in [-15, 15]$; true effect 0 for $k < 0$ and 1 for $k \geq 0$ (dashed). Navy line: mean across replications; grey band: 90% Monte Carlo band; thin grey lines: six example replications. Pre-trends are exactly parallel by construction while post-period paths are dominated by $\Delta_\lambda F_t$.

Here $V(k, \hat{F}^k)$ is the squared loss

$$V(k, \hat{F}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \lambda_{k,i} F_{kt})^2,$$

evaluated at factors estimated by principal components, since the true F_t is unobserved. Two features of the construction matter in practice. First, the variance estimate $\hat{\sigma}^2$ entering the PC penalty is taken from the model with $k_{\max}$ factors rather than from the model being evaluated, so that every component of the penalty other than k itself is held fixed as k varies. Second, $V(k, \hat{F}^k)$ equals the sum of all but the first k eigenvalues of $X'X/(NT)$, which places these criteria in the same eigenvalue framework as the estimators described below and makes the comparison across the four approaches transparent.

The three penalty functions suggested by [Bai and Ng \(2002\)](#), combined with the two criterion types, give six statistics:

$$\begin{aligned} \text{PC}_{p1} &= V(k, \hat{F}^k) + k \hat{\sigma}_{k_{\max}}^2 \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right) \\ \text{PC}_{p2} &= V(k, \hat{F}^k) + k \hat{\sigma}_{k_{\max}}^2 \left(\frac{N+T}{NT} \right) \ln \{\min(N, T)\} \\ \text{PC}_{p3} &= V(k, \hat{F}^k) + k \hat{\sigma}_{k_{\max}}^2 \left(\frac{\ln \{\min(N, T)\}}{\min(N, T)} \right) \\ \text{IC}_{p1} &= \ln V(k, \hat{F}^k) + k \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right) \\ \text{IC}_{p2} &= \ln V(k, \hat{F}^k) + k \left(\frac{N+T}{NT} \right) \ln \{\min(N, T)\} \\ \text{IC}_{p3} &= \ln V(k, \hat{F}^k) + k \left(\frac{\ln \{\min(N, T)\}}{\min(N, T)} \right). \end{aligned}$$

Each statistic is evaluated over $k \in \{1, 2, \dots, k_{\max}\}$ and the estimated number of factors is the value of k that *minimises* it. This is worth signalling because the two criteria described next are maximisers rather than minimisers.

[Ahn and Horenstein \(2013\)](#) suggest two estimators to estimate the number of common factors F : the eigenvalue ratio (ER) and growth rate (GR) estimators. The estimators for a given number of common factors k

$$\begin{aligned}\text{ER}(k) &= \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}}, \quad k = 1, 2, \dots, k_{\max} \\ \text{GR}(k) &= \frac{\ln(1 + \tilde{\mu}_{NT,k}^*)}{\ln(1 + \tilde{\mu}_{NT,k+1}^*)}\end{aligned}$$

with $\tilde{\mu}_{NT,k}$ being the k th largest eigenvalue of $X'X/(NT)$ and $\tilde{\mu}_{NT,k}^* = \tilde{\mu}_{NT,k}/V(k)$, where $V(k)$ is the mean of the squared residuals of a regression of X on the first k principal components of $X'X/(NT)$. Here X is the $T \times N$ matrix collecting the residualized data, so that $X'X/(NT)$ is $N \times N$; since the non-zero eigenvalues of $X'X/(NT)$ and $XX'/(NT)$ coincide, either orientation may be used and both criteria are computed over $k \leq \min(N, T)$. The number of common factors is then the maximiser, not the maximum,

$$\begin{aligned}\tilde{k}_{\text{ER}} &= \arg \max_{1 \leq k \leq k_{\max}} \text{ER}(k) \\ \tilde{k}_{\text{GR}} &= \arg \max_{1 \leq k \leq k_{\max}} \text{GR}(k).\end{aligned}$$

Onatski (2010) proposes identifying the number of factors by examining the gaps between the ordered eigenvalues of $X'X/T$. He notes that eigenvalues beyond the true number of factors tend to cluster around a finite value, creating a threshold δ . The method selects the smallest eigenvalue among the top $k_{\max}$ whose gap with the next smaller eigenvalue exceeds δ , as this indicates the presence of a genuine factor.

The following algorithm is used:

1. Obtain the eigenvalues $\mu_1, \mu_2, \dots, \mu_N$ of $T^{-1} \sum_{t=1}^T X_t X_t'$. Set $j = k_{\max} + 1$.
2. Estimate the slope coefficient $\hat{\beta}$ in a linear regression of $\mu_j, \dots, \mu_{j+4}$ on $(j - 1)^{2/3}, \dots, (j + 3)^{2/3}$ and a constant, and set the threshold δ to $\delta = 2|\hat{\beta}|$.
3. Obtain the estimated number of factors as $\hat{r}(\delta) = \max(i \leq r_{\max}^N : \mu_i - \mu_{i+1} \geq \delta)$ if there exists some $i \leq r_{\max}^N$ subject to $\mu_i - \mu_{i+1} \geq \delta$. Otherwise, set $\hat{r}(\delta) = 0$.
4. Set $j = \hat{r}(\delta)$. Repeat steps 2 and 3 until $\hat{r}(\delta)$ remains unchanged in two subsequent repetitions. Let this final estimator be denoted $\tilde{r}_{\text{ED}}$.

Gagliardini et al. (2019) propose a postestimation method based on residuals from linear models to distinguish between weak cross-sectional dependence and a factor structure. The approach compares eigenvalues of the residual covariance matrix to a penalty term $g(N, T)$. The penalty is positive by construction, so what signals a factor is not its sign but whether an eigenvalue exceeds it: at least one factor is present when the largest eigenvalue lies above

the penalty, that is when $\xi(1) > 0$. The number of factors is then the number of eigenvalues that clear the penalty.

The difference between the k -th largest eigenvalue, $\mu(k)$, and the penalty $g(N, T)$ is

$$\begin{aligned}\xi(k) &= \mu(k) - g(N, T) \\ g(N, T) &= \frac{(\sqrt{N} + \sqrt{T})^2}{NT} \ln \left\{ \frac{NT}{(\sqrt{N} + \sqrt{T})^2} \right\}\end{aligned}\tag{30}$$

The number of factors is then selected according to

- no factors: $\xi(1) < 0$
- otherwise $\hat{k} = \min\{k \in \{1, \dots, T-1\} : \xi(k) < 0\} - 1$,

that is, one less than the first index at which the difference turns negative, so that $\hat{k}$ counts the eigenvalues lying above the penalty. Note that $\xi(k)$ is defined for $k \geq 1$, the k -th largest eigenvalue, so the index set starts at one; the no-factor case is the $\hat{k} = 0$ instance of the same rule. If $\xi(k) > 0$ for all $k = 1, \dots, T-1$, then $\hat{k} = T$.

OA.2 Online Appendix: Complements to the Application

OA.2.1 Time Series of PLS by Country and Sub-Sector

Each figure plots the quarterly PLS series for every sub-sector of one country over 2002–2024.

FIGURE OA.2: Time Series of PLS by Sub-Sector: Austria

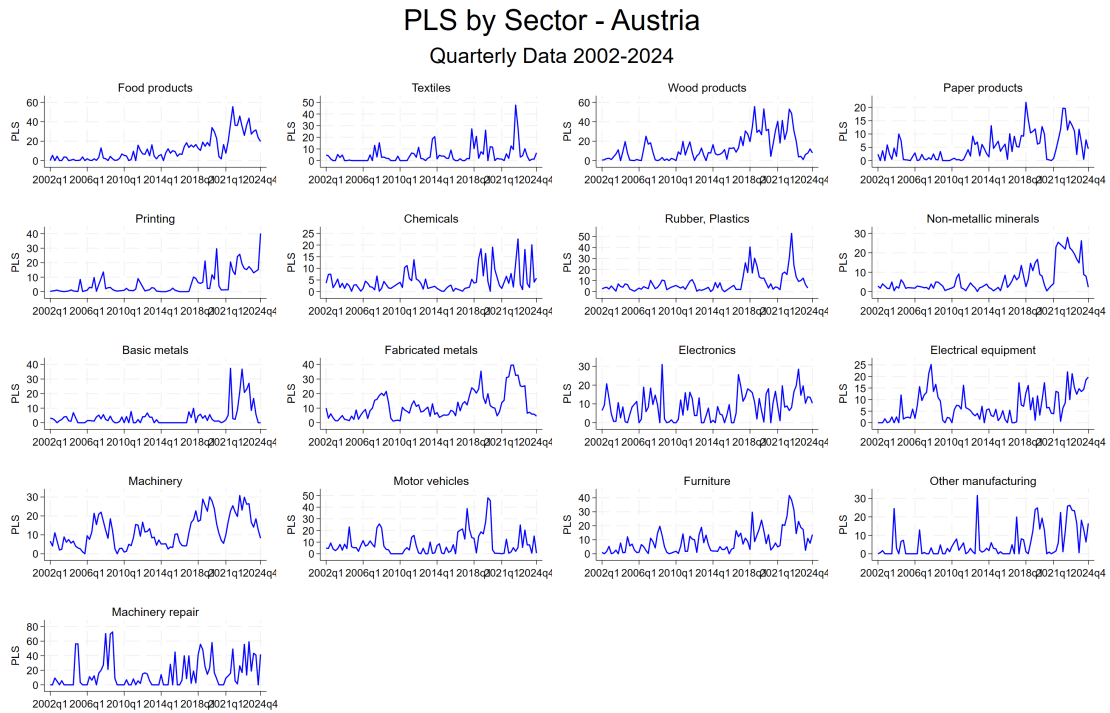


FIGURE OA.3: Time Series of PLS by Sub-Sector: Belgium

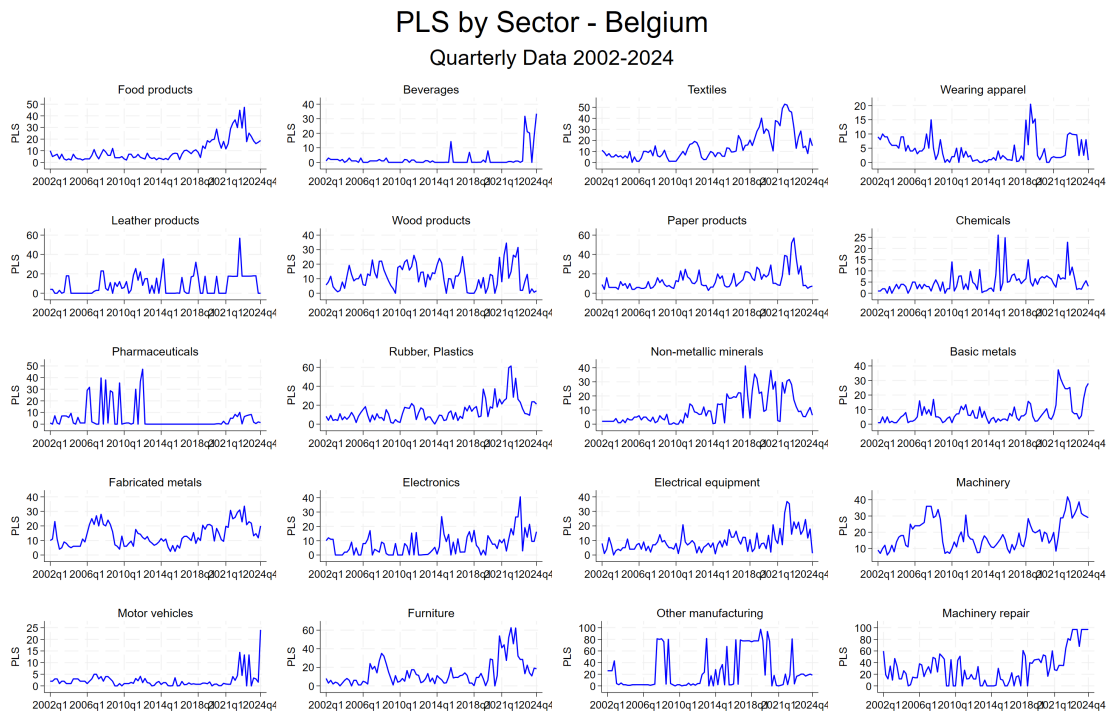


FIGURE OA.4: Time Series of PLS by Sub-Sector: Germany

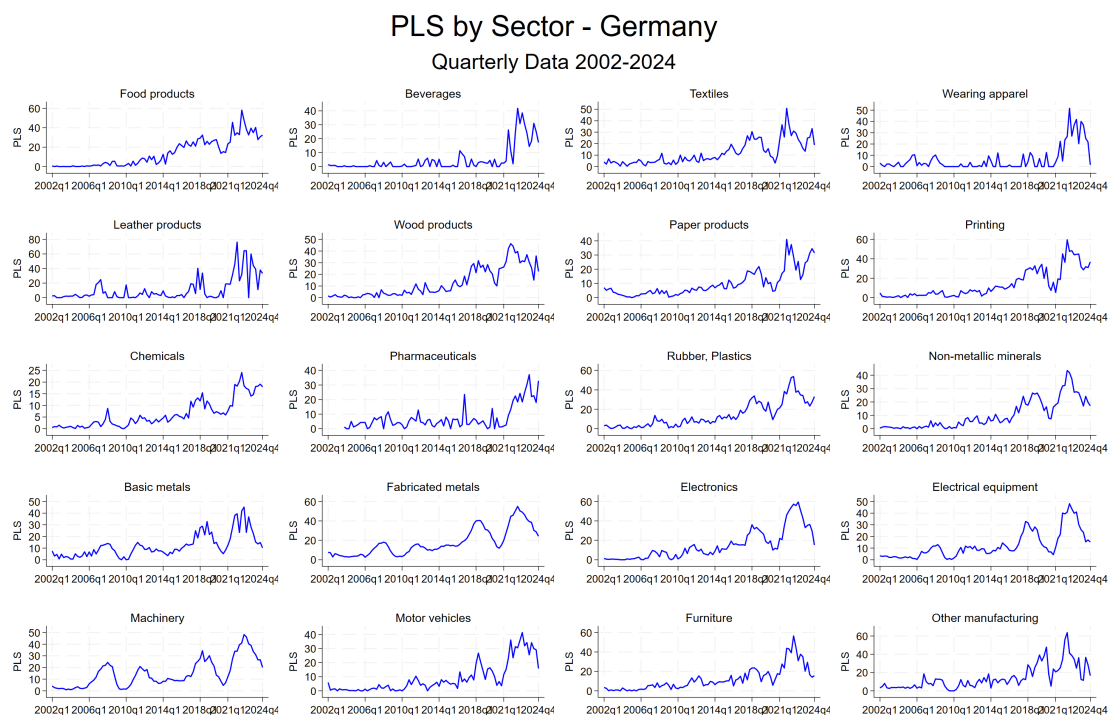


FIGURE OA.5: Time Series of PLS by Sub-Sector: Greece

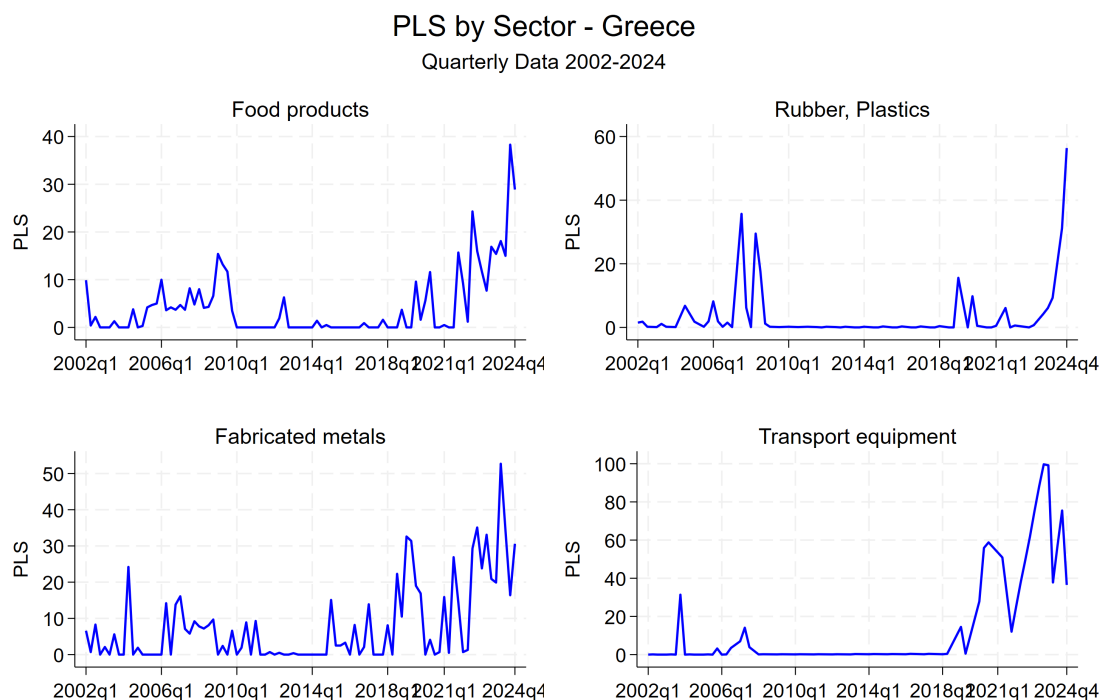


FIGURE OA.6: Time Series of PLS by Sub-Sector: Spain

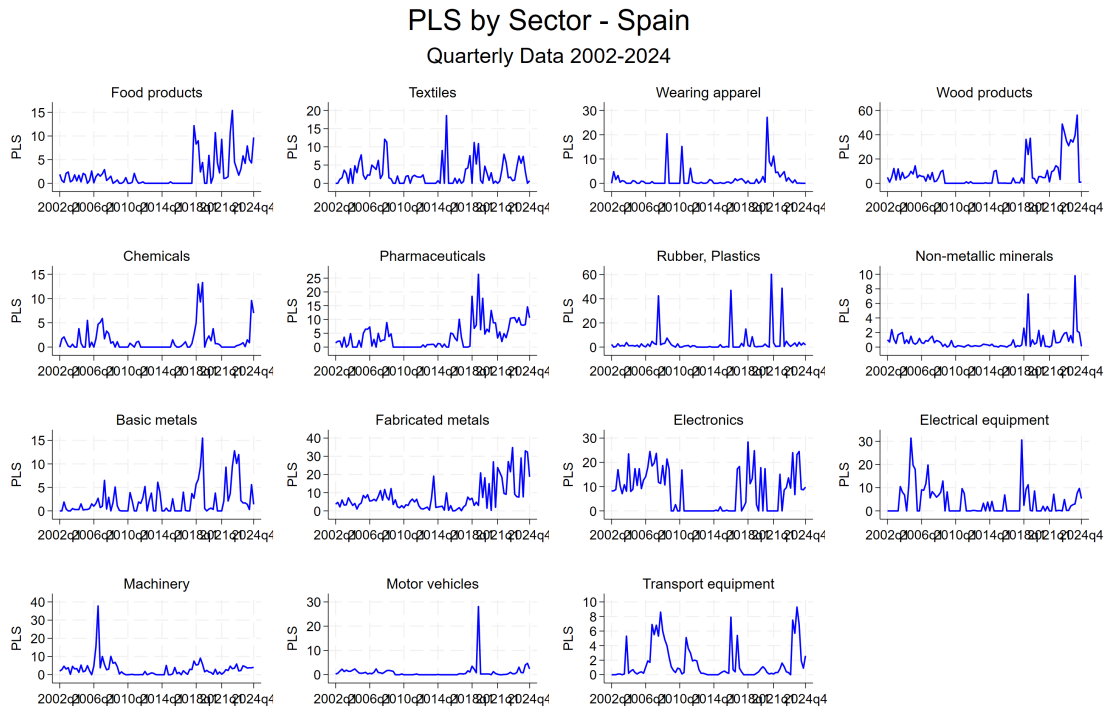


FIGURE OA.7: Time Series of PLS by Sub-Sector: Finland

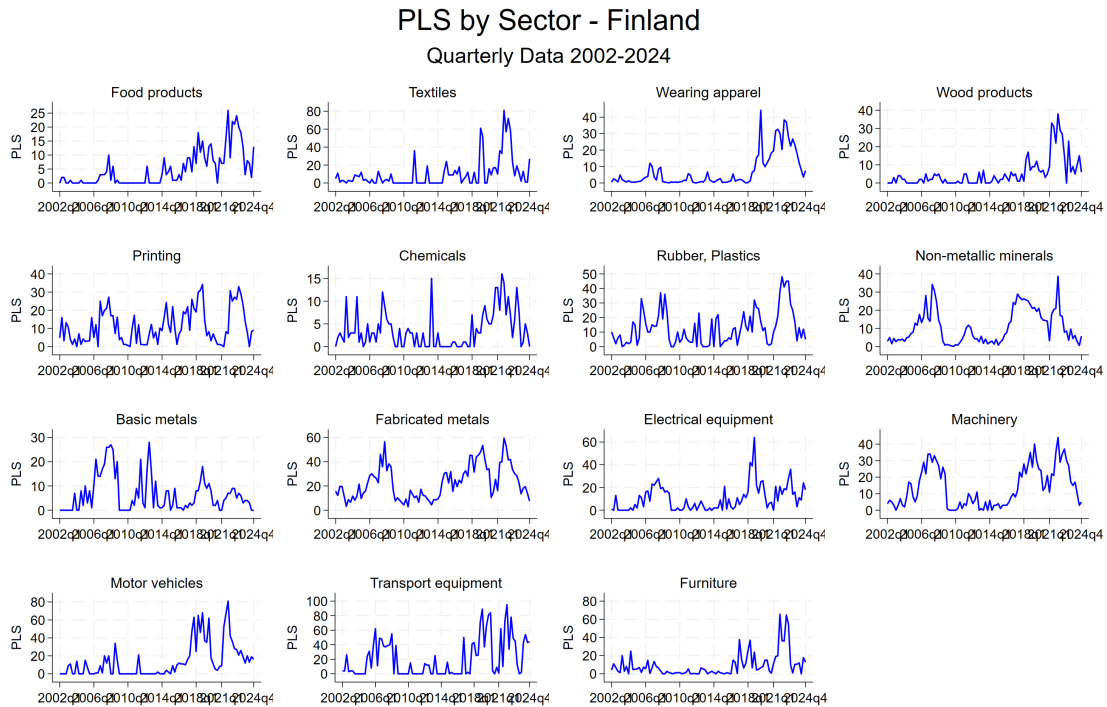


FIGURE OA.8: Time Series of PLS by Sub-Sector: France

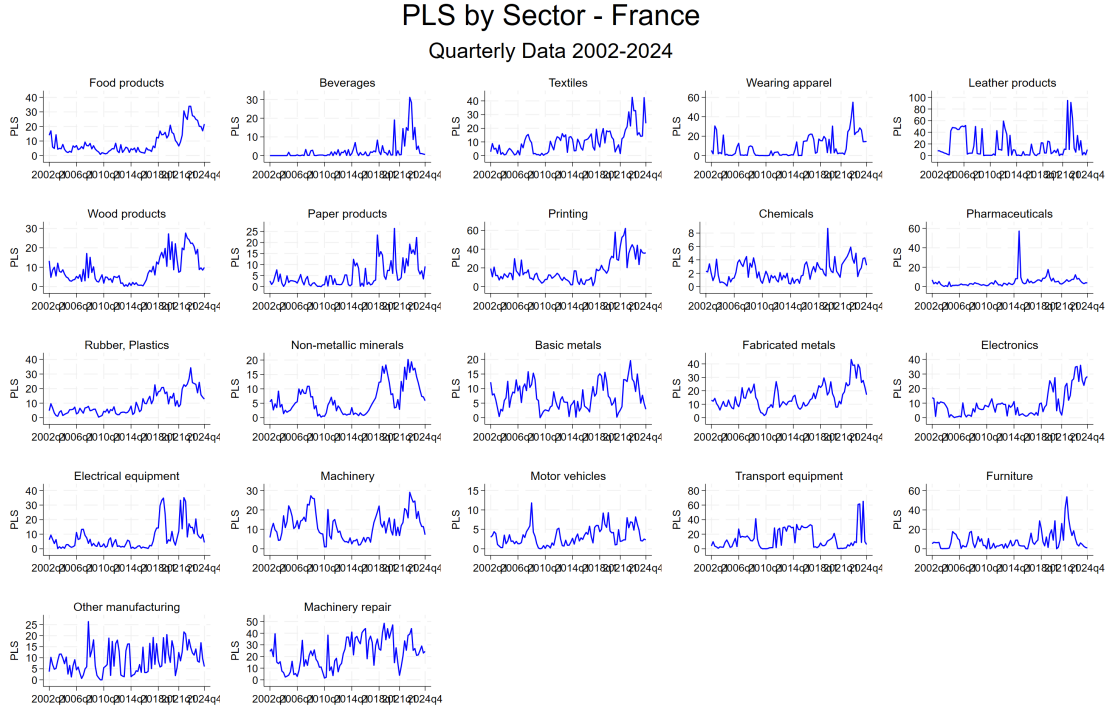


FIGURE OA.9: Time Series of PLS by Sub-Sector: Italy

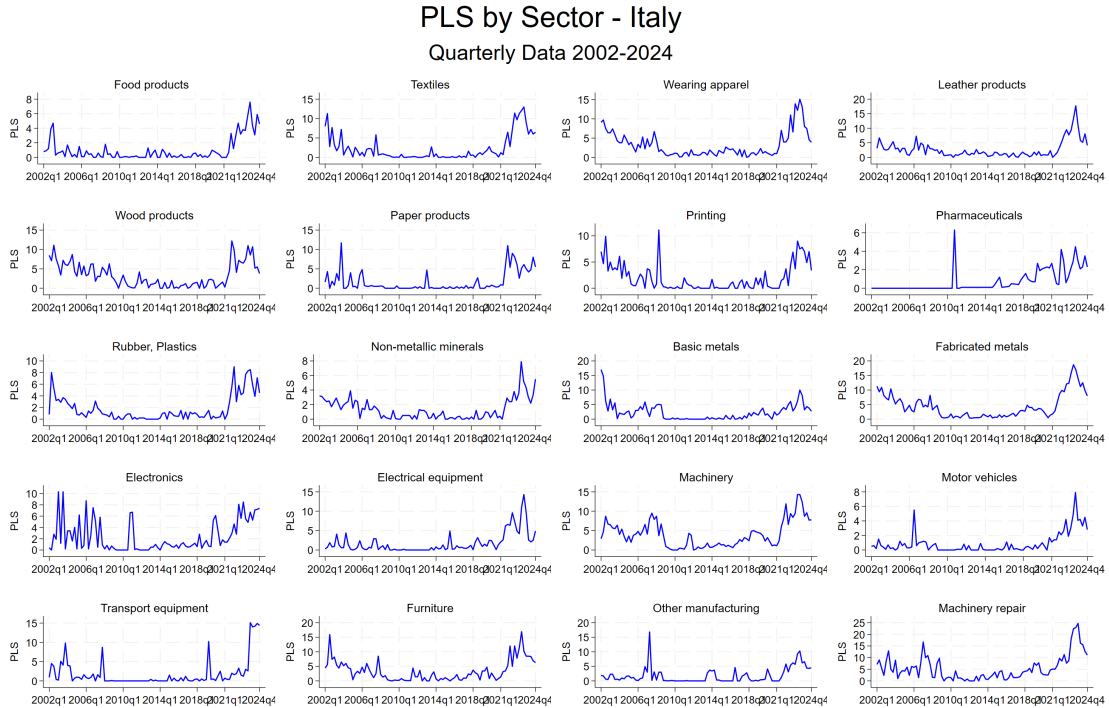


FIGURE OA.10: Time Series of PLS by Sub-Sector: the Netherlands

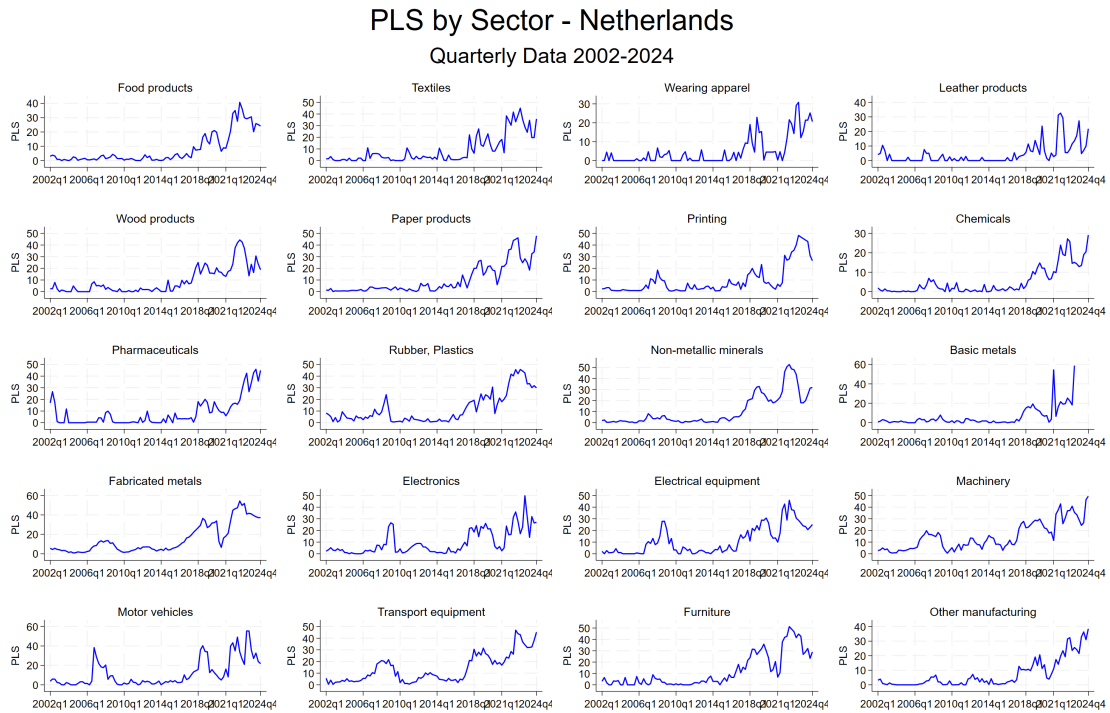


FIGURE OA.11: Time Series of PLS by Sub-Sector: Portugal

