Online Appendix

“Measuring Trend Inflation Based on Financial Market Conditions: The Case of Japan” by Masahiko Shibamoto

A1 Data sources

- Consumer price index for all items (2020=100); consumption-tax-adjusted series for the periods from April 1997 to March 1998 and April 2014 to March 2015; calculated retroactively for the period preceding December 1989 using the monthly change in the index and adding 1.2 in March 1989 to eliminate the influence of the consumption tax instituted from April 1989; retrieved from the Ministry of Internal Affairs and Communications.
- 10-year Japanese government bond yields (end of month, %); retrieved from NIKKEI NEEDS FINANCIAL QUEST.
- M2 money stock; Seasonally adjusted series (monthly average, 100 million yen); retrieved from the Bank of Japan.
- Nominal effective exchange rates by the BIS; 2020 AVERAGE=100.
- CPI for all items excluding fresh foods, all items excluding fresh food and energy (2020=100); consumption-tax-adjusted series; retrieved from the Ministry of Internal Affairs and Communications.
- CPI Trimmed mean, Weighted median, and Mode; retrieved from the Bank of Japan.

A2 Testing for the existence of a long-run relationship between inflation and financial market conditions

This online appendix reports the test results on the existence of a long-run relationship between inflation and financial market conditions. Specifically, we report the results of unit root tests on the series representing inflation and financial market conditions, and cointegration tests on the interdependence between inflation and financial market conditions.

A2.1 Unit root tests

Table A1 shows the results of unit root tests for series for inflation and financial market conditions. For the autoregressive models in the test, we consider a model that includes only a constant and a model that includes both a constant and a time trend. The lag order included in each test model are chosen by general to specific reduction (from fifteen lags) proposed by Hall (1994). ADF represents the Augmented Dickey and Fuller (1979) (ADF) test statistic under the null hypothesis that the series follows a unit root process; KPSS represents the Kwiatkowski et al. (1992) (KPSS) test statistic under the null hypothesis that the series follows a stationary process.

Table A1: Unit root test results

Null hypothesis	ADF Unit root	Unit root	KPSS Stationary	Stationary
	Constant	Constant and trend	Constant	Constant and trend
In levels				
Inflation	-2.87	-2.47	1.053	0.424
10-year yield	-1.99	-2.00	4.342	0.729
Money stock (log)	-1.81	-3.26	4.619	0.701
Exchange rate (log)	-3.01	-2.72	2.406	0.643
In first differences				
Inflation	-6.75	-6.90	0.174	0.046
10-year yield	-5.77	-5.92	0.194	0.034
Money stock (log)	-2.56	-2.77	1.278	0.544
Exchange rate (log)	-5.05	-5.27	0.325	0.034

		10%	5%	1%
ADF	Constant	-2.59	-2.87	-3.46
	Constant and trend	-3.16	-3.43	-4.00
KPSS	Constant	0.347	0.463	0.739
	Constant and trend	0.119	0.146	0.216

Note: ADF denotes the Augmented Dickey and Fuller (1979) (ADF) test (null: unit root). KPSS denotes the Kwiatkowski et al. (1992) (KPSS) test (null: stationarity). For each series, we report results from specifications with (i) a constant and (ii) a constant plus a linear time trend. The lag order in each auxiliary regression is selected using the general-to-specific procedure of Hall (1994), starting from 15 lags. Critical values are from Said and Dickey (1984) for ADF and Kwiatkowski et al. (1992) for KPSS. Sample period: February 1983-December 2022.

Table A1 reports ADF and KPSS unit-root tests for inflation and the financial variables in levels and in first differences. In levels, the KPSS test strongly rejects stationarity for all series, consistent with the view that inflation and the financial variables are highly persistent and can be treated as nonstationary over the sample. In first differences, the evidence is broadly consistent with stationarity for most series. These results suggest that the series representing inflation and series for financial market conditions follow a non-stationary process.

A caveat concerns the order of integration of the money stock. While we treat the log money stock as an I(1) variable in the baseline analysis, unit-root tests applied to its first difference yield results that the ADF test fails to reject the unit-root null, whereas the KPSS test rejects the stationarity null. This suggests that the integration order of money stock may be difficult to pin down in finite samples and we therefore cannot fully rule out higher-order integration or I(2) behavior. To ensure that our conclusions are not driven by the time-series assumption of the money stock as an I(1), we also conduct robustness checks using alternative dynamic specifications that allows for the possibility that the log money stock exhibits I(2) behavior.

A2.2 Cointegration tests

Next, we test for the existence of a cointegration relationship between inflation and a series representing financial market conditions. Specifically, we perform the Engle and Granger (1987) (EG) test, the Johansen (1995) (JOH) test, and the Choi et al. (2008) (CHO) test.

Table A2 shows the results of the cointegration test on the interdependence between inflation and financial market conditions. The analysis provides statistical support for the existence of a long-run relationship between inflation and financial market conditions. According to the EG test results, the null hypothesis that the cointegration residual follows a unit root process is rejected at a significance level of 1%. According to the JOH test results, the null hypothesis that there is no cointegration relationship is rejected at the 1% significance level, and the null hypothesis that there is at most one cointegration relationship is not rejected at the normal significance level. Furthermore, the CHO test results show that we cannot reject the null hypothesis that a cointegration relationship exists at the usual significance level. These results imply that there is one common stochastic trend between inflation and financial market conditions.

Table A2: Cointegration tests between inflation and financial conditions.

	EG	JOH		CHO
Null hypothesis	No coint	# of coint ≤ 0	# of coint ≤ 1	Coint
Test statistics	-5.30	57.77	25.90	3.99 [0.26]

		10%	5%	1%
EG		-3.92	-4.22	-4.88
JOH	# of coint≤ 0	44.49	47.86	54.68
	# of coint≤ 1	27.07	29.80	35.46

Note: This table reports cointegration tests for the long-run relationship in equation (1) in the main text. EG is the Engle and Granger (1987) residual-based ADF test (13 lag order selected by the Hall (1994) general-to-specific procedure from 15 lags). JOH is the Johansen (1995) trace test (VAR lag order = 13). Critical values are from Phillips and Ouliaris (1990) for EG and Mackinnon et al. (1999) for JOH. CHO is the Hausman-type cointegration test of Choi et al. (2008). Bracketed values are p-values computed from a chi-square distribution with 3 degrees of freedom. Sample period: February 1983-December 2022.

A3 Structural stability of the cointegrating relationship

A concern with long samples is that the cointegrating relationship may not be stable. We therefore test parameter stability in the cointegrating regression using the SupF and MeanF tests of Hansen (1992), where the alternative allows for an unknown break date. Table A3 reports the test statistics and the corresponding critical values. Both tests fail to reject stability at conventional significance levels, supporting the use of recursive (expanding-window) estimation when constructing the real-time indicator.

Table A3: Stability test results for the cointegrating relationship.

	SupF	MeanF
Test statistics	10.6	1.88

	10%	5%	1%
SupF	15.3	17.2	21.0
MeanF	6.66	7.68	10.10

Note: This table reports Hansen's (1992) tests for parameter stability in the cointegrating regression. SupF tests the null of constant parameters against a one-time structural break at an unknown date. MeanF tests the null of parameter constancy against the alternative of gradual parameter variation (random-walk-type instability). For both tests, the Bartlett kernel is used with a truncation lag of 5. Critical values are from Hansen (1992).

Although these tests do not indicate parameter instability at conventional levels, allowing for time-varying parameters is a natural extension to study gradual changes in the long-run relationship.

A4 Uncertainty of recursive estimates

To quantify uncertainty in the real-time trend inflation estimates, we compute pointwise confidence bands for $\pi^{rme}_{t|t}$. For each date t, we re-estimate the DOLS regression in equation (2) in the main text using data up to t and obtain the covariance matrix of the estimated long-run coefficients. Let $x_t = (1, r_t, m_t, e_t)'$ and $\Phi_{|t} = (\phi_{1|t}, \phi_{r|t}, \phi_{m|t}, \phi_{e|t})'$. Then

$$\widehat{\mathrm{Var}}(\pi^{rme}_{t|t}) = \widehat{\mathrm{Var}}(x_t' \Phi_{|t}) = x_t' \, \widehat{\mathrm{Var}}(\Phi_{|t}) \, x_t,$$

where $\widehat{\mathrm{Var}}(\Phi_{|t})$ is computed using Newey and West (1987) heteroskedasticity and autocorrelation robust standard errors. Figure A1 reports 90% pointwise bands $(\pi^{rme}_{t|t} \pm 1.65\sqrt{\widehat{\mathrm{Var}}(\pi^{rme}_{t|t})})$, showing that the confidence intervals remain reasonably tight even in earlier parts of the recursive sample. This alleviates concerns that limited sample size could make the real-time indicator too imprecise for practical use.

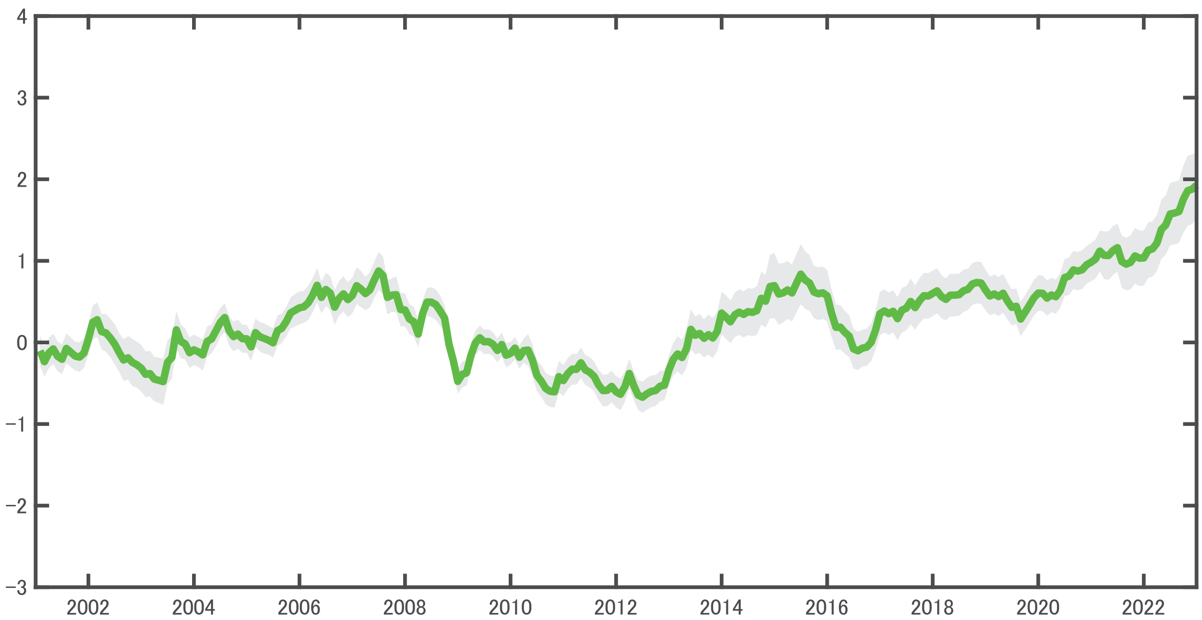

Figure A1: Recursive common-trend estimates with 90% confidence bands.

Notes: The sample period runs from January 2001 to December 2022. The solid line shows the recursive estimate of the common trend. Shaded areas denote 90% pointwise confidence bands (±1.65 standard errors).

A5 Robustness of recursive trend inflation

This section reports robustness checks corresponding to "3.3 Additional analyses" discussed in the main text. The results are summarized in Figures A2-A4 and Table A4.

A5.1 Alternative cointegration estimator: FMOLS

First, we replace DOLS with the fully modified OLS (FMOLS) estimator (Phillips and Hansen, 1990), which provides an alternative correction for endogeneity in cointegrated systems. Figure A2 compares the resulting real-time common-trend estimates, and Table A4 reports the out-of-sample prediction comparison. The qualitative conclusions are unchanged: the common-trend forecast remains competitive in 2001-2012 and performs best in 2013-2021, and the Model Confidence Set results continue to favor the common trend in the later period.

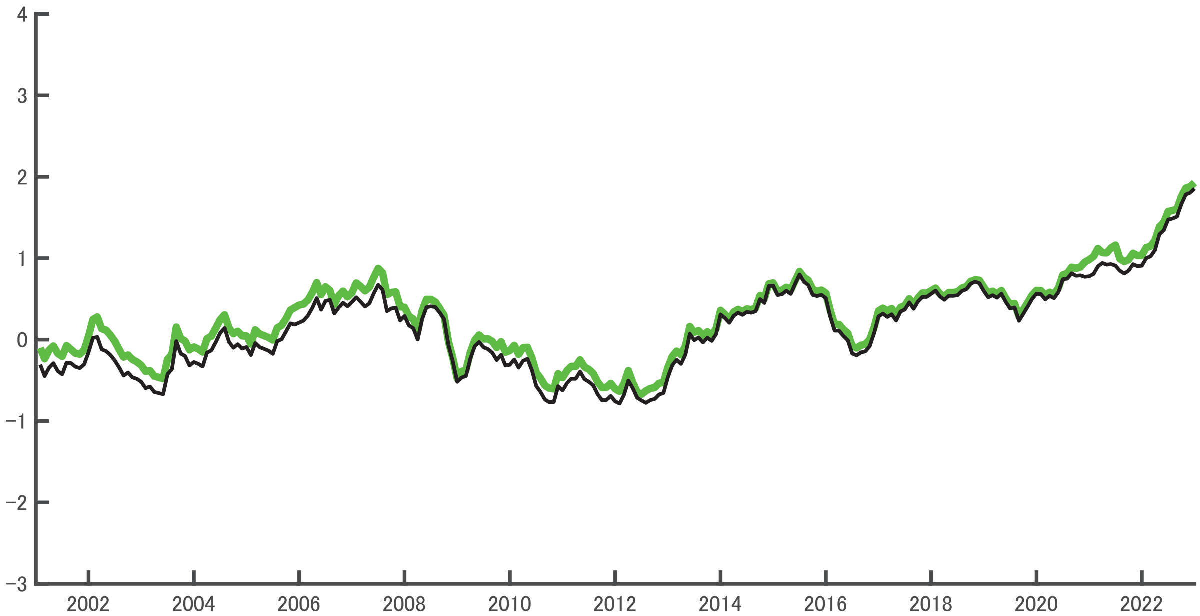

Figure A2: Recursive common-trend estimates: benchmark vs. FMOLS.

Notes: The thin line shows the recursive common-trend estimate obtained using fully modified OLS (FMOLS). The long-run covariance matrix is estimated with a Bartlett kernel and a truncation bandwidth set to $T_t^{1/4}$, where T_t is the recursive sample size at time t. The bold line shows the benchmark recursive estimate obtained using DOLS. The sample period shown is January 2001-December 2022.

A5.2 Sensitivity to the starting date of the real-time exercise

Second, we assess sensitivity to the starting date used for recursive estimation. Because recursive estimation accumulates information over time, early sample choices could in principle matter

Table A4: Robustness checks for prediction comparison

	Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prediction period: January 2001 - December 2012								
MSE	FMOLS	0.48	1.07	0.94	0.60	0.33	0.40	0.52
	Start:Feb-1980	0.55						0.53
	Start:Jun-1985	0.48						0.52
	Start:Nov-1986	0.57						0.53
	$m_t \sim$ I(2)	0.47						0.52
DM	FMOLS	-	-1.77	-3.00	-0.82	1.95	0.87	-0.41
	Start:Feb-1980	-	-1.53	-2.33	-0.34	2.26	1.37	0.16
	Start:Jun-1985	-	-1.81	-3.07	-0.87	2.19	1.03	-0.45
	Start:Nov-1986	-	-1.45	-2.18	-0.20	2.31	1.45	0.27
	$m_t \sim$ I(2)	-	-1.82	-3.08	-0.92	1.98	0.84	-0.58
MCS p-value	FMOLS	0.173	0.042	0.000	0.129	1.00	0.173	0.173
	Start:Feb-1980	0.247	0.052	0.000	0.219	1.00	0.247	0.247
	Start:Jun-1985	0.143	0.043	0.000	0.122	1.00	0.143	0.143
	Start:Nov-1986	0.264	0.054	0.000	0.264	1.00	0.264	0.264
	$m_t \sim$ I(2)	0.162	0.042	0.000	0.123	1.00	0.162	0.123
Prediction period: January 2013-December 2021								
MSE	FMOLS	0.89	2.15	2.23	1.43	1.39	1.27	1.43
	Start:Feb-1980	0.65						1.38
	Start:Jun-1985	0.80						1.43
	Start:Nov-1986	0.82						1.43
	$m_t \sim$ I(2)	0.86						1.44
DM	FMOLS	-	-3.06	-2.45	-2.23	-1.97	-1.62	-2.19
	Start:Feb-1980	-	-3.01	-2.41	-2.37	-2.05	-1.82	-2.21
	Start:Jun-1985	-	-3.05	-2.46	-2.31	-2.04	-1.75	-2.24
	Start:Nov-1986	-	-3.04	-2.44	-2.28	-2.01	-1.71	-2.21
	$m_t \sim$ I(2)	-	-3.08	-2.49	-2.27	-2.06	-1.72	-2.27
MCS p-value	FMOLS	1.00	0.003	0.010	0.084	0.084	0.084	0.012
	Start:Feb-1980	1.00	0.002	0.013	0.027	0.027	0.031	0.016
	Start:June-1985	1.00	0.003	0.013	0.052	0.052	0.052	0.013
	Start:Nov-1986	1.00	0.003	0.013	0.057	0.057	0.057	0.014
	$m_t \sim$ I(2)	1.00	0.003	0.012	0.075	0.075	0.075	0.012

(1) Common trend (2) all items excluding fresh food (3) all items excluding fresh food and energy (4) Trimmed mean (5) Weighted median (6) Mode (7) Combination (Average (1)-(6))

Note: MSE denotes the mean-squared prediction error over two prediction periods, January 2001 to December 2012 and January 2013 to December 2021. DM denotes the Diebold and Mariano (1995) statistic for testing the significance of predictive superiority of common trend forecast with respect to each core inflation forecast. All DM tests are implemented with Newey and West (1987) heteroskedasticity and a serial correlation robust covariance matrix with the truncation parameter $P^{1/4}$ where P is the prediction sample size. MCS p-values report the inclusion p-value p_{MCS} from the Model Confidence Set (MCS) procedure of Hansen et al. (2011). The loss is squared forecast error for 24-month-ahead average inflation. p-values are obtained by a block bootstrap (5000 replications; block length $P^{1/4}$).

for initial parameter estimates. To avoid arbitrariness, we select alternative starting dates that coincide with business-cycle turning points (peaks or troughs) in the official business-cycle chronology. This design ensures that the initial estimation window begins at economically meaningful dates and spans different cyclical phases. Figure A3 reports common-trend estimates obtained under alternative starting dates in addition to the benchmark. Table A4 shows that the forecast performance of the common-trend indicator and its relative ranking are robust to these alternatives.

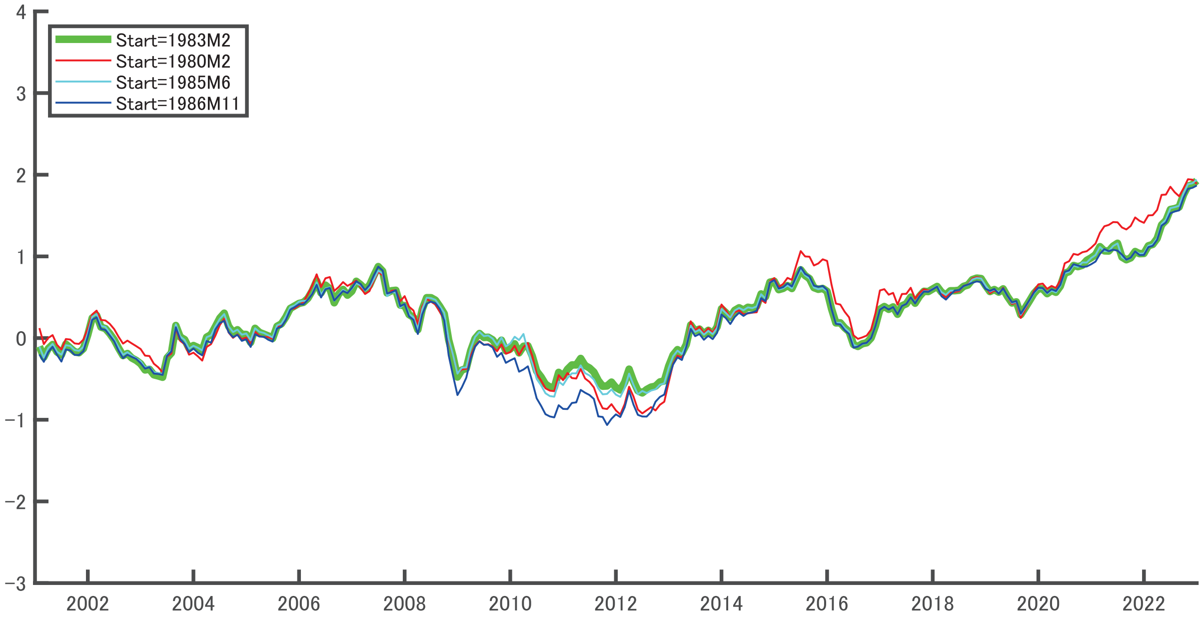

Figure A3: Recursive common-trend estimates using alternative starting dates.

Notes: Each thin line shows the recursive common-trend estimate using an alternative starting date (February 1980, June 1985, or November 1986). The bold line shows the benchmark recursive estimate (starting date: February 1983). The sample period shown is January 2001-December 2022.

A5.3 Allowing for possible I(2) behavior in money stock

Third, motivated by the evidence on the order of integration for money in first differences, we allow for a conservative I(2) specification for money stock. Specifically, we re-estimate the DOLS regression by including leads and lags of $\Delta^2 m_t$ (instead of Δm_t) as control variables while maintaining the cointegrating relation among (π_t, r_t, m_t, e_t). Figure A4 compares the resulting common-trend estimates to the benchmark. Table A4 shows that the forecasting results remain close to the benchmark, indicating that the main findings are not driven by the exact treatment of money dynamics in the short-run controls.

Figure A4: Recursive common-trend estimates allowing for possible I(2) behavior in money stock.

Notes: The thin line shows the recursive common-trend estimate using an augmented DOLS specification that includes leads and lags of $\Delta r_t, \Delta^2 m_t$, and Δe_t to allow for possible I(2) behavior in the money stock. The bold line shows the benchmark recursive estimate. The sample period shown is January 2001-December 2022.

A6 Sensitivity analyses

A6.1 Rolling-window estimation of trend inflation

We additionally report rolling-window (fixed-length) estimation as a sensitivity check. We implement 18-year rolling windows (196 months) and update the cointegrating regression and the implied trend estimate each month.

Figure A5 compares the rolling-window trend estimates with the benchmark recursive estimate. The rolling-window estimates can be unstable not only because the window is shorter, but because fixed-length windows may contain insufficient low-frequency variation to identify the stochastic trend accurately. This issue is particularly relevant for Japan, where inflation exhibited relatively little low-frequency movement from the early 2000s to the early 2010s.

This instability is visible not only in the implied trend series but also in the estimated cointegrating parameters. Figure A6 plots the rolling-window estimates of the long-run coefficients $(\phi_{r|t}, \phi_{m|t}, \phi_{e|t}, \phi_{1|t})$ over time. The coefficient paths display noticeable swings–sometimes changing markedly from one window to the next-particularly in periods when inflation exhibits little low-frequency movement (e.g., from the early 2000s to the early 2010s). In such windows,

Figure A5: Common-trend estimates: benchmark (recursive) vs. rolling-window estimation.

Notes: The thin line shows the rolling-window estimate of the common trend (window length: 18 years = 196 months). The bold line shows the benchmark recursive estimate. The sample period shown is January 2001-December 2022.

the long-run relationship is less precisely pinned down because the data contain limited information about low-frequency comovements. As a result, the rolling-window estimates become more sensitive to small changes in the sample composition, and the resulting trend series inherits this instability. By contrast, recursive (expanding-window) estimation accumulates low-frequency information over time and yields substantially more stable parameter updates.

For these reasons, we treat rolling-window results as an auxiliary exercise and rely on recursive estimation as the main approach.

A6.2 Year-over-year transformations of financial variables

One would suggest using year-over-year changes of financial variables as robustness checks. We therefore replace (r_t, m_t, e_t) with their 12-month changes and re-run the cointegration analysis and estimation. However, this transformation is not well aligned with the objective of extracting a common stochastic trend from the levels of inflation and financial conditions. By construction, 12-month differencing removes low-frequency movements, which are precisely the component the cointegration approach is designed to identify.

In this case, the exercise effectively estimates a pseudo-trend driven by a misspecified long-

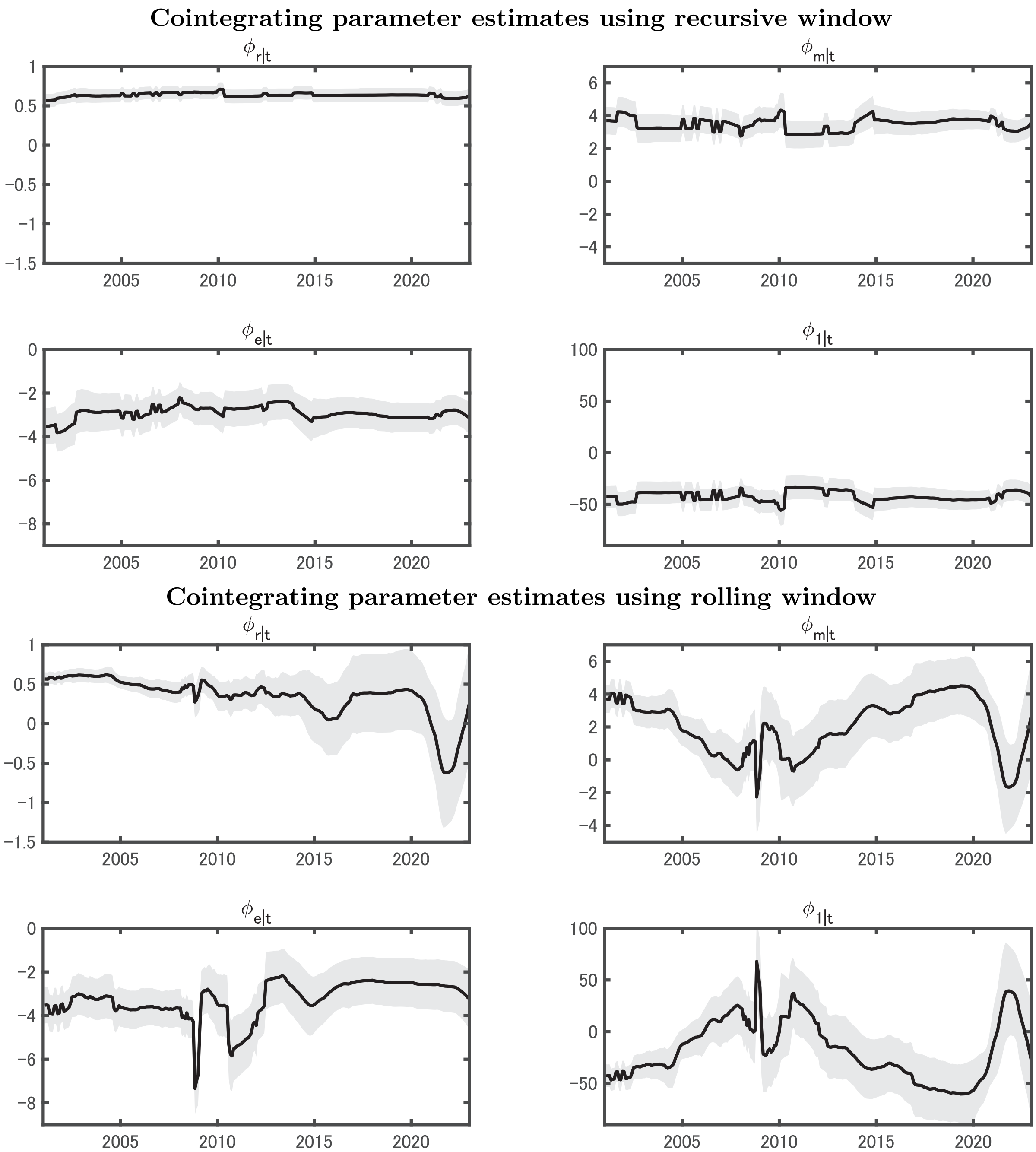

Figure A6: Updated cointegrating-parameter estimates: recursive vs. rolling windows.

Notes: Upper panels show recursive estimates of each cointegrating coefficient, and lower panels show rolling-window estimates. Solid lines show estimates of each cointegrating coefficient. Shaded areas denote 90% pointwise confidence bands for each estimate. The sample period shown is January 2001-December 2022.

run relation rather than by a common stochastic trend. More importantly, the transformation undermines the time-series basis for a trend interpretation. As shown in Table A5, standard cointegration tests provide weak support for a stable cointegrating relationship under year-over-year transformations. When cointegration is not supported, regressions in levels can behave like *spurious* relationships, and the implied "trend inflation" constructed from such estimates is no longer a coherent estimate of an underlying stochastic trend. Consistent with this concern, Table A6 shows that long-run coefficients under the year-over-year specification differ markedly from the benchmark (levels) specification and the estimated long-run coefficients become highly sensitive to the estimation method: DOLS and Dynamic Feasible Generalized Least Squares (DFGLS) yield markedly different results.[1] This suggests that the implied "trend" is not a coherent estimate of a common stochastic trend and can behave in a *spurious-regression*-like manner. Figure A7 illustrates that the resulting series behaves in a way that is difficult to interpret as trend inflation.

For these reasons, we treat the year-over-year transformations of financial variables as an auxiliary exercise rather than a meaningful robustness check for our trend-inflation objective. The benchmark level-based specification, supported by the cointegration evidence and stability tests, remains the appropriate framework for estimating the common trend.

Table A5: Cointegration tests between inflation and year-over-year changes in financial conditions.

	EG	JOH		CHO
Null hypothesis	No coint	# of coint ≤ 0	# of coint ≤ 1	Coint
Test statistics	-3.18	74.72	43.41	6.77 [0.08]

		10%	5%	1%
EG		-3.92	-4.22	-4.88
JOH	# of coint≤ 0	44.49	47.86	54.68
	# of coint≤ 1	27.07	29.80	35.46

Note: This table reports cointegration tests between inflation and year-over-year (12-month) changes in financial conditions. EG is the Engle and Granger (1987) residual-based ADF test, JOH is the Johansen (1995) trace test, and CHO is the Hausman-type test of Choi et al. (2008). Bracketed values are p-values computed from a chi-square distribution with 3 degrees of freedom. Sample period: February 1983-December 2022. See the note to Table A2 for details on lag choices and critical values.

[1]Choi et al. (2008) propose the DFGLS estimator, which is robust whether or not the error in the regression is stationary or unit-root nonstationary.

Table A6: Long-run coefficient estimates: levels (benchmark) vs. year-over-year transformations.

	r_t, m_t, e_t: levels (benchmark)		r_t, m_t, e_t: year-over-year changes	
	DOLS	DFGLS	DOLS	DFGLS
$\phi_{r\|T}$	0.64	0.60	-0.28	0.02
	(0.05)	(0.11)	(0.24)	(0.27)
$\phi_{m\|T}$	3.57	3.29	16.65	7.36
	(0.42)	(0.69)	(3.63)	(5.58)
$\phi_{e\|T}$	-3.12	-2.89	-0.37	-0.06
	(0.45)	(0.78)	(1.41)	(1.59)

Note: This table reports estimates of the long-run coefficients in equation (2) in the main text. The benchmark specification uses the levels of financial variables as proxies for (r_t, m_t, e_t), while the alternative specification uses their year-over-year (12-month) changes. DOLS and DFGLS denote dynamic ordinary least squares and dynamic feasible generalized least squares, respectively. Lead and lag orders for differenced regressors are selected by the Akaike Information Criterion. Numbers in parentheses are Newey and West (1987) heteroskedasticity and autocorrelation robust standard errors for least squares with 5-month lag truncation. Sample period: February 1983 to December 2022.

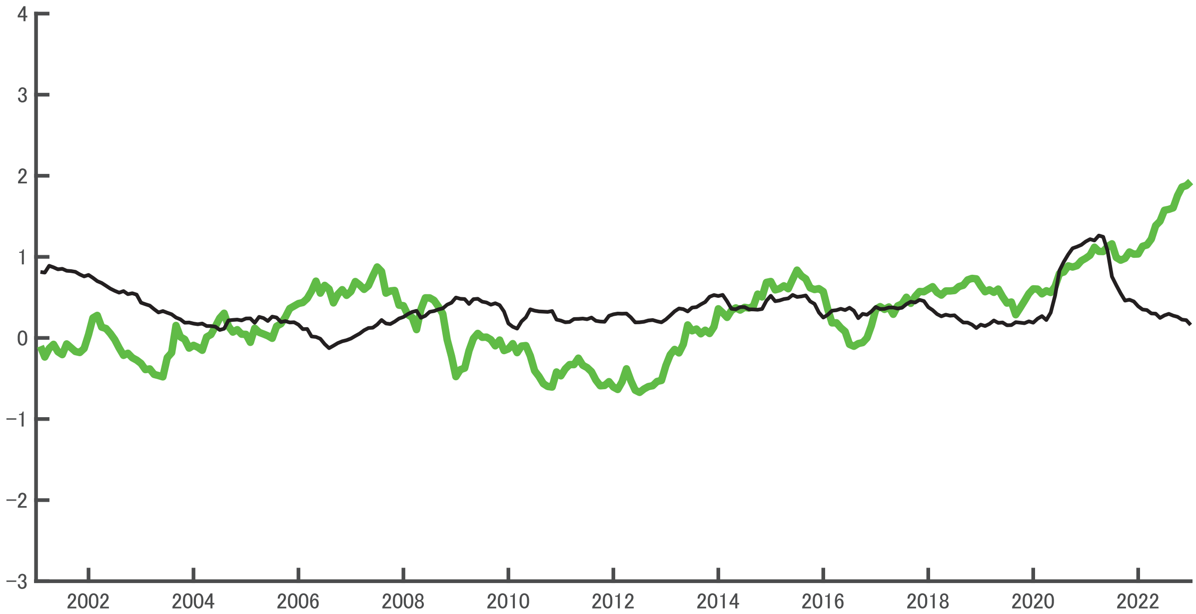

Figure A7: Common-trend estimates: benchmark vs. year-over-year transformations of financial variables.

Notes: The thin line shows the common-trend estimate constructed using year-over-year (12-month) changes in financial variables. The bold line shows the benchmark common-trend estimate in levels. The sample period shown is January 2001-December 2022.

References

Choi, Chi-Young, Ling Hu, and Masao Ogaki. 2008. "Robust Estimation for Structural Spurious Regressions and a Hausman-type Cointegration Test." *Journal of Econometrics* 142 327–351.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." *Journal of the American Statistical Association* 74 (366): 427–431.

Diebold, Francis, and Roberto Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13 (3): 253–263.

Engle, Robert F., and C. W. J. Granger. 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55 (2): 251–276.

Hall, Alastair. 1994. "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection." *Journal of Business and Economic Statistics* 12 (4): 461–470.

Hansen, Bruce E. 1992. "Tests for Parameter Instability in Regressions with I(1) Processes." *Journal of Business and Economic Statistics* 10 (3): 321–335.

Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. "The Model Confidence Set." *Econometrica* 79 (2): 453–497.

Johansen, Soren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54 (1-3): 159–178.

Mackinnon, James G., Alfred A. Haug, and Leo Michelis. 1999. "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration." *Journal of Applied Econometrics* 14 563–577.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708.

Phillips, Peter C. B., and Bruce E. Hansen. 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes." *The Review of Economic Studies* 57 (1): 99–125.

Phillips, Peter C. B., and Sam Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58 (1): 165–193.

Said, Said E., and David A. Dickey. 1984. "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order." *Biometrika* 71 (3): 599–607.