

When Does Monetary Policy Sway House Prices? A Meta-Analysis *

Dominika Ehrenbergerova^{a,b}, Josef Bajzik^{a,b}, Tomas Havranek^{b,c}

^aCzech National Bank, Na Prikope 28, 115 03 Prague, Czech Republic

^bIES FSV Charles University, Opletalova 26, 110 00 Prague, Czech Republic

^cCEPR, London, United Kingdom

September 2022

forthcoming in *IMF Economic Review*

Abstract

Several central banks have leaned against the wind in the housing market by increasing the policy rate preemptively to prevent a bubble. Yet the empirical literature provides mixed results on the impact of short-term interest rates on house prices: the estimated semi-elasticities range from -12 to positive values. To assign a pattern to these differences, we collect 1,555 estimates from 37 individual studies that cover 45 countries and 72 years. We then relate the estimates to 39 characteristics of the financial system, business cycle, and estimation approach. Our main results are threefold. First, the mean reported estimate is exaggerated by publication bias, because insignificant results are underreported. Second, inclusion of controls correlated with policy rates (credit or money supply) decreases the estimated effects of policy rates on house prices. Third, the effects are stronger in countries with more developed mortgage markets and generally later in the cycle when the yield curve is flat and house prices enter an upward spiral.

Keywords: Interest rates, house prices, monetary policy transmission, meta-analysis, publication bias, Bayesian model averaging

JEL Codes: C83, E52, R21

*An online appendix with data and code is available at meta-analysis.cz/house_prices.

Dominika Ehrenbergerova, dominika.ehrenbergerova@cnb.cz, corresponding author.

Josef Bajzik, josef.bajzik@cnb.cz.

Tomas Havranek, tomas.havranek@ies-prague.org.

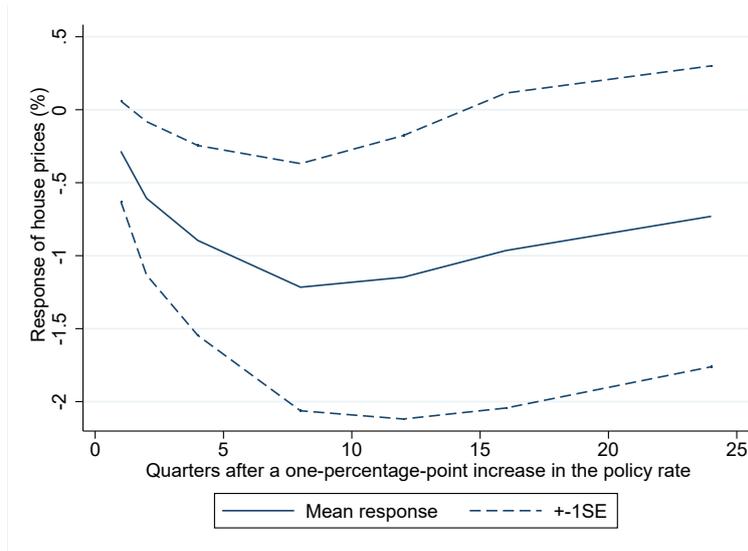
The opinions expressed here are ours and not necessarily those of the Czech National Bank.

We thank the editor and two anonymous referees for their excellent comments. We also thank Simona Malovana, Marek Rusnak, Jakub Mateju, and seminar participants at the Czech National Bank and MAER-Net Colloquium for their helpful comments on an earlier version of this paper. Bajzik and Ehrenbergerova acknowledge support from the Czech Science Foundation (project 21-09231S). Havranek acknowledges support from the Czech Science Foundation (project 19-26812X). This work was supported by the NPO Systemic Risk Institute (project LX22NPO5101).

1 Introduction

Common wisdom has it that monetary policy is largely responsible for asset bubbles, including the rising house prices. That view sometimes translates into policy, such as in the case of the Swedish Riksbank between 2010 and 2014 or the government of New Zealand in 2021. In the most famous example of leaning against the wind, the Riksbank increased its policy rate from near zero to 2% in order to tame household indebtedness and house prices, even at substantial costs in terms of inflation and unemployment (Svensson, 2014, 2017). The government of New Zealand, in turn, recently amended the mandate of the Reserve Bank of New Zealand and instructed it to consider house prices when making monetary policy decisions (Powell & Wessel, 2021). The policy change in New Zealand is interesting both because its Reserve Bank has been an influential pioneer of innovations in central banking (introducing inflation targeting in 1990) and because by 2021 a large amount of research has amassed on the effects of monetary policy on house prices. This recent research, however, is rarely cited in the policy debate, which remains influenced by the arguments of Taylor (2007) in favor of the effectiveness of short-term rates in taming bubbles.¹ Perhaps one of the reasons for the relatively limited impact of the recent research is the variance in results. The literature lacks a synthesis that would assign a pattern to the different conclusions. That is what we attempt to provide in this paper.

Figure 1: Mean reported response of house prices to a monetary tightening



Notes: Computed based on 1,555 estimates from 237 impulse responses reported in 37 papers. On average, the response bottoms out after two years at a 1.2% decrease in house prices following a one-percentage-point increase in the policy rate.

Figure 1 shows the mean response of house prices to a one-percentage-point increase in the short-term monetary policy rate. The mean is extracted from 237 impulse responses reported in 37 studies. The impulse responses, computed from vector autoregressions (VARs, Sims, 1980), are the main output of these studies. Hence our meta-analysis is unusual in that we collect and examine graphical results: the exact numerical results are rarely reported. For selected time

¹For an excellent recent discussion of the pros and cons of leaning against the wind, see Benati (2021).

horizons after the monetary policy shock we measure pixel coordinates and collect the estimated response of house prices. Meta-analyses of graphical results are rare, and a prominent recent example is the meticulous survey by Fabo *et al.* (2021) on the effects of quantitative easing. Note that Figure 1 shows the corresponding 68% confidence interval (one standard error on both sides of the mean), which is the norm in the VAR literature. (Few impulse responses would be statistically significant at the 5% level common in most other fields of economics.) The impulse response bottoms out after two years at a 1.2% decrease in house prices following a one-percentage-point increase in the policy rate. We will call this effect, here 1.2, a semi-elasticity. It is clear that, on average, with such a small semi-elasticity central banks cannot plausibly combat double-digit inflation in house prices. The implication echoes Williams (2016), a concise narrative survey of 10 earlier papers on the topic.

But a mean figure conceals important differences in the context in which the impulse response is estimated. Perhaps in some countries and certain phases of the business cycle, leaning against the wind can help moderate the increase in house prices (and, vice versa, a loose policy may help reflate depressed housing markets). Calza *et al.* (2013) suggest that the transmission of monetary policy to house prices is stronger in countries with larger flexibility and development of mortgage markets. Similarly, Iacoviello & Minetti (2003) show that financial liberalization can be important for the strength of transmission. Assenmacher-Wesche & Gerlach (2010) examine whether transmission differs between boom and standard periods. Or perhaps the small mean response is contaminated by measurement problems, such as simple recursive identification (a problem stressed, for example, by Bjørnland & Jacobsen, 2010) and omission of important variables, such as credit (Assenmacher-Wesche & Gerlach, 2010). Our comparative advantage to the studies mentioned above is the richness of the meta-analysis dataset. No previous study in this literature has used data for more than 19 countries, which has made it difficult to investigate cross-country differences. Few cross-country studies examine more than a couple of business cycles. Similarly, comparisons of results with different identification of VAR models within individual studies have so far lacked statistical power. The work of the researchers who have collectively produced 237 impulse responses for various contexts allows us to examine the heterogeneity in transmission systematically.

Another problem with the mean impulse response is potential publication bias (Stanley, 2001),² which stems from the selective reporting of results that have the intuitive sign or are statistically significant. Vector autoregressions are complex models with (at least in this literature) typically few degrees of freedom. It follows that the resulting impulse response is sometimes counterintuitive: for example, it can show that house prices do not react to policy rates, or even more puzzlingly that house prices rise following a monetary tightening. If researchers take such results as evidence that their model is misspecified, they can try to run

²For recent papers on publication bias in economics, including positive and negative evidence, see Havranek (2015), Brodeur *et al.* (2016), Bruns & Ioannidis (2016), Ioannidis *et al.* (2017), Stanley & Doucouliagos (2017), Stanley *et al.* (2017), Card *et al.* (2018), Christensen & Miguel (2018), Astakhov *et al.* (2019), DellaVigna *et al.* (2019), Blanco-Perez & Brodeur (2020), Brodeur *et al.* (2020), Ugur *et al.* (2020), Xue *et al.* (2020), Imai *et al.* (2021), Neisser (2021), Brown *et al.* (2022), Iwasaki (2022), and DellaVigna & Linos (2022). Earlier influential papers include Card & Krueger (1995), Stanley (2005, 2008), and Stanley & Doucouliagos (2010).

different specifications until they obtain the desired outcome. The problem is that while the puzzling impulse responses can indeed arise because of misspecifications, they can also appear simply by chance, especially given the small datasets in the literature. Seemingly large estimated effects of monetary policy in the right direction can also be due to misspecifications or chance, but it is difficult to identify them. Zero is a clear psychological cutoff that is not mirrored by a corresponding upper threshold and thereby causes a bias towards larger effects. The resulting publication bias can only be corrected using meta-analysis techniques.

Publication bias does not imply cheating and is inevitable in observational empirical research even if all researchers are honest. (In experimental research the bias can potentially be tackled by the preregistration of experiments, see, for example, Olken, 2015, but preregistration is difficult when data are publicly available, so that the researcher can inspect them before preregistration). Publication selection can even improve the results of individual studies. The underlying effect of policy rate hikes on house prices will most likely be negative in most if not all contexts, so it is likely that the “wrong” sign indeed suggests to a researcher a problem with specification, sample size, or both. Thus it will improve the conclusions of an individual study when it does not focus on positive or zero responses of house prices. The idea of sign restrictions in vector autoregressions, eloquently advocated by Uhlig (2005), builds on a related principle. Unfortunately, under selective reporting the literature becomes biased as a whole since large estimates, also given by chance or misspecifications, are rarely omitted. So with individual studies we never know how much they suffer from publication bias.

For the basic identification of publication bias correction techniques we use the analogy suggested by McCloskey & Ziliak (2019), who compare publication selection to the Lombard effect in psychoacoustics: speakers involuntarily increase their vocal effort with increasing background noise. Similarly, given the example in the previous paragraph, many researchers will try harder to change the specification of their vector autoregression model if they have small samples and thus a lot of noise in estimation, a noise that often leads to insignificant initial estimates. With sufficient effort, the VAR model can be adjusted in a way that produces point estimates large enough to outweigh the large standard errors and thus delivers statistical significance. Therefore, selective reporting creates a correlation between estimates and standard errors, a correlation that otherwise should not appear in the literature. Aside from linear tests based on the Lombard effect (regressions of estimates on standard errors) we also employ recently developed nonlinear techniques by Andrews & Kasy (2019), Furukawa (2021), and van Aert & van Assen (2021). The latter technique, p-uniform*, relaxes the assumption of no correlation between estimates and standard errors in the absence of publication bias; the assumption is perhaps too strong for the VAR literature where the impulse responses are nonlinear combinations of underlying (unreported) regression coefficients. All techniques agree that the exaggeration due to publication bias is at least twofold.

In the second part of the analysis we relate the estimated impulse responses to the context in which they were obtained. To this end we collect 39 variables that reflect the characteristics of data (e.g. time coverage), specification (e.g. inclusion of long-term interest rates), estima-

tion (e.g. nonrecursive identification), publication (e.g. the number of citations per year), and countries (e.g. the mean share of mortgages with a floating rate in the period for which the impulse response was estimated). To tackle model uncertainty in relating the estimated semi-elasticities to the 39 explanatory variables we employ Bayesian (Raftery *et al.*, 1997; Eicher *et al.*, 2011; Steel, 2020) and frequentist (Hansen, 2007; Amini & Parmeter, 2012) model averaging. We address collinearity by using the dilution prior (George, 2010). The finding of substantial publication bias is robust to controlling for heterogeneity. Regarding data characteristics, our results suggest that studies covering shorter time series tend to produce stronger responses of house prices to monetary shocks (that is, larger semi-elasticities in the absolute value), which is consistent with a small-sample bias. Regarding specification characteristics, we find that the omission of variables related to liquidity (credit or money supply) is associated with stronger responses of house prices to changes in the policy rate. The effect is substantial and can strengthen the reported semi-elasticity by one percentage point. In contrast, we find little evidence that estimation and publication characteristics help explain the heterogeneity observed in the literature.

The factors most useful in explaining the differences in impulse responses are variables reflecting structural heterogeneity: the characteristics of the countries and periods for which the impulse responses were produced. Three variables are especially important. First, it is the degree of development of the mortgage market (and credit markets in general). With larger credit markets in relation to GDP, the transmission of monetary policy to house prices gets stronger. Second, it is the slope of the yield curve. With flatter yield curves, the reported semi-elasticities are larger in the absolute value. Third, it is the period of a prolonged rise in house prices: when house prices have increased for several years, monetary policy becomes more potent at taming them. These country- and time-level characteristics can alter the implied impulse response by up to three percentage points. Therefore while on average house prices do not respond much to monetary policy, policy rates can help alleviate the build-up of housing bubbles in countries with developed mortgage markets during the latter part of the business cycle. Such alleviation is nevertheless costly in terms of inflation and unemployment, because even the most optimistic estimates implied by our analysis for outlying countries and time periods suggests that, after correction for publication bias, a one-percentage-point increase in the policy rate is associated with a maximum decrease in house prices of less than 3%.

2 The Semi-Elasticity Dataset

We collect estimates of the effect of changes in the policy rate on house prices. In general, these estimates are produced in the modern literature by two types of models: dynamic stochastic general equilibrium (DSGE) models and vector autoregression (VAR) models. The results of both can be interpreted as empirical estimates, though always conditional on theoretical considerations. DSGE models need to be calibrated (or their priors set), and of course their structure is entirely based on theory. The identification of VAR models, in turn, often has theoretical foundations as well, but in some cases only as an afterthought. Compared to DSGE,

VAR models are generally more data-driven, and the corresponding estimates are thus better suited for meta-analysis methods. Moreover, DSGE estimates of the semi-elasticity are relatively rare (a prominent example being Iacoviello & Neri, 2010).³ To avoid comparing apples and oranges, we focus on VAR estimates only. A general structural VAR model has the following form:

$$A_0^i Y_t^i = a^i + \gamma^i t + A^i(L) Y_{t-1}^i + B^i(L) z_t + e_t^i, \quad (1)$$

where Y_t^i is a vector of endogenous variables (including policy rates and house prices) for time t and country i , a^i is a constant, $A^i(L)$ and $B^i(L)$ are distributed lag polynomials, z_t is a vector of exogenous variables, and e_t^i is an error term. The set of endogenous variables in a relevant VAR model usually includes output in addition to short-term rates and house prices. Depending on model specification, it may also include other variables, such the exchange rate, consumption, money supply, long-term interest rates, residential investment, and credit. In order to estimate (1), researchers rewrite it in reduced form. The principal outputs from VAR models, the reactions of the endogenous variables to structural shocks, are usually reported graphically as impulse response functions, which are easy for the reader to interpret and which cover the response over several time horizons.

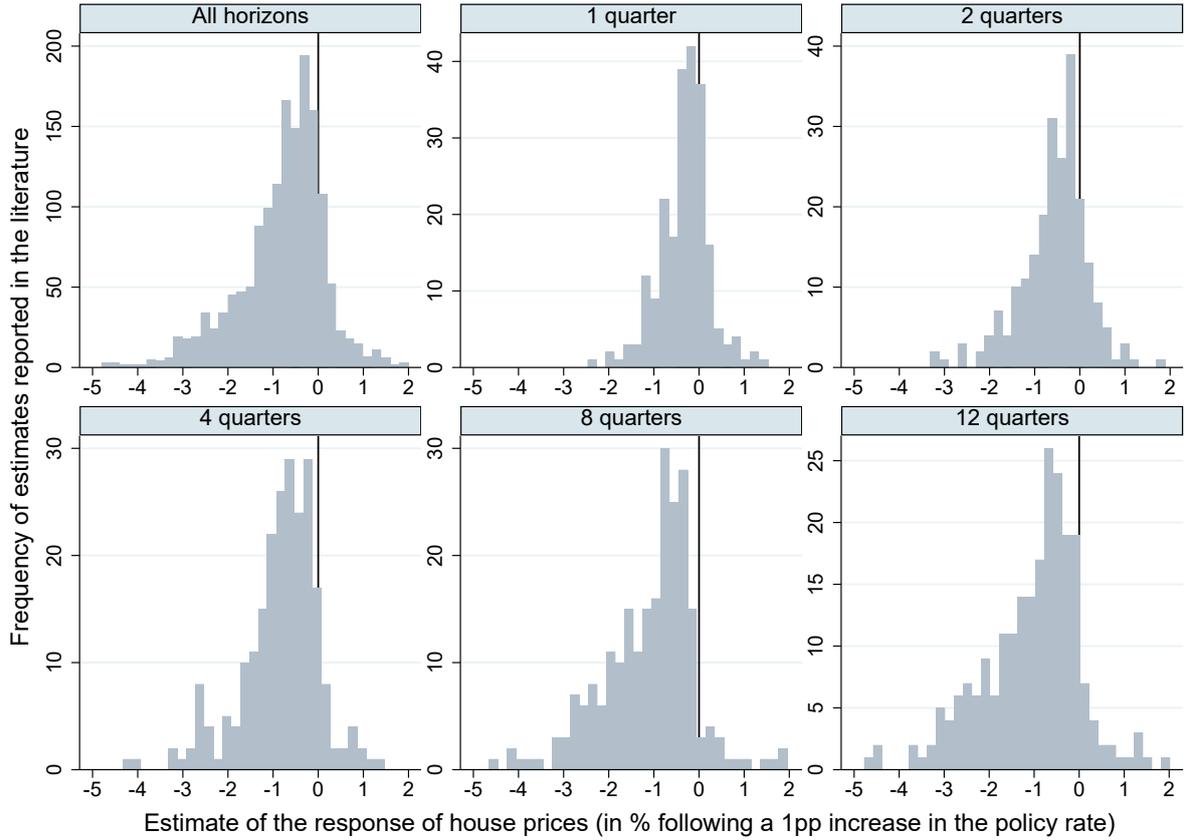
To search for relevant studies we use Google Scholar because of its broad coverage and full-text capabilities. (More details on our search strategy, including the exact query, are available in Appendix A.) We calibrate our search query in order to obtain the best known studies among the first hits. We inspect the first 500 papers produced by the search. Following Fabo *et al.* (2021), we also inspect the references and citations of the studies identified by the baseline search. Each study needs to fulfill the following three criteria: First, for quantitative comparability the study must use a VAR model that includes house prices (not house price inflation); we thus cannot use a few influential studies such as Fratantoni & Schuh (2003) and Del Negro & Otrok (2007).⁴ Second, monetary policy stance must be proxied by the short-term interest rate. Third, the study must report confidence intervals around the impulse response function so that we can recover the precision of the estimate, which is essential for tests of publication bias. These criteria leave a total of 37 studies, both published articles and working papers (all written in English), which collectively use unbalanced data from 45 countries between 1947 and 2018. We add the last study in April 2022. The list of included studies is available in Table A1 in Appendix A.

From these 37 studies we collect the responses of house prices to a change in the policy rate after one, two, four, eight, twelve, sixteen, and twenty-four quarters. In each case we measure pixel coordinates using WebPlotDigitizer to recover the numerical estimate as precisely as possible. The extracted values are checked for consistency with authors' verbal assessment

³There are only 5 DSGE studies (together providing only 7 impulse responses) that satisfy our inclusion criteria (e.g., report confidence intervals for their impulse responses).

⁴Estimates from 8 studies on house price inflation can be converted to level estimates using the approach outlined in Fabo *et al.* (2021). But because we focus on publication bias, we also need standard errors for these estimates. The standard errors can be approximated only roughly using the delta method, which is why we include the corresponding studies merely as a robustness check in the online appendix (Table B3). The results do not change qualitatively.

Figure 2: Reported effects of monetary policy on house prices at different horizons



Notes: Outliers are omitted from the graphs for ease of exposition but included in all statistical tests.

of their own findings. In the online appendix we provide an archive of all impulse responses together with the extracted numerical values. We gather 222 and 227 responses after one and two quarters, respectively, and label these as short-term effects. To capture mid-term effects we gather 237 estimates for both the four- and eight-quarter horizons. To capture long-term effects we collect 232 estimates for the twelve-quarter horizon, 226 estimates for the sixteen-quarter horizon, and 174 estimates for the twenty-four-quarter horizon. Because many studies do not report responses at the latter horizon, in the analysis we focus on horizons up to sixteen quarters, and in particular the mid-term effects (four and eight quarters) most relevant to monetary policy. In a few cases, the responses for the short-term effects (one and two quarters) are not reported as the corresponding impulse responses start at the four-quarter horizon. For each impulse response we standardize the effects so that they correspond to a percentage response of house prices to a one-percentage-point increase in the policy rate. We compute the standard error from the reported confidence intervals; in the few cases when the confidence intervals are asymmetrical, we approximate the standard errors by taking the average of both bounds.

We have already commented on the mean impulse response function, Figure 1, in the Introduction. A closer view of the distribution of semi-elasticities at different horizons is provided in Figure 2. At the one-quarter horizon, most of the estimates are close to zero, and the dis-

tribution is almost symmetrical. With an increasing horizon, the mass of the estimates moves to the left, and the distribution becomes asymmetrical. Note that very few estimates suggest a large response of house prices to changes in the policy rate. A couple of outliers are cut from the figure for ease of exposition (the largest one being -12%), but these are isolated cases. In total for all the horizons, 87% of all the semi-elasticities lie between -2% and 1% . Moreover, more than 50% of all the semi-elasticities lie between -1% and 0% .

In addition to the impulse response functions, we also collect 39 control variables that capture the specifics of each study in order to examine the heterogeneity in the estimates. Slightly fewer than two thirds of the variables included are collected from primary studies themselves, while the remaining third consist of external country-level variables included to examine structural heterogeneity and collected from the World Bank, OECD, and Eurostat. In accordance with the latest meta-analysis reporting guidelines (Havranek *et al.*, 2020), the data taken from individual studies (estimates, confidence intervals, and variables reflecting estimation context) were collected by two co-authors of this paper and cross-checked to eliminate potential mistakes arising from manual collection. These variables are discussed in more detail in Section 4, which focuses on the heterogeneity in the literature. In the next section we focus on publication bias, which can distort the reported semi-elasticities shown in Figure 1 and Figure 2.

3 Publication Bias

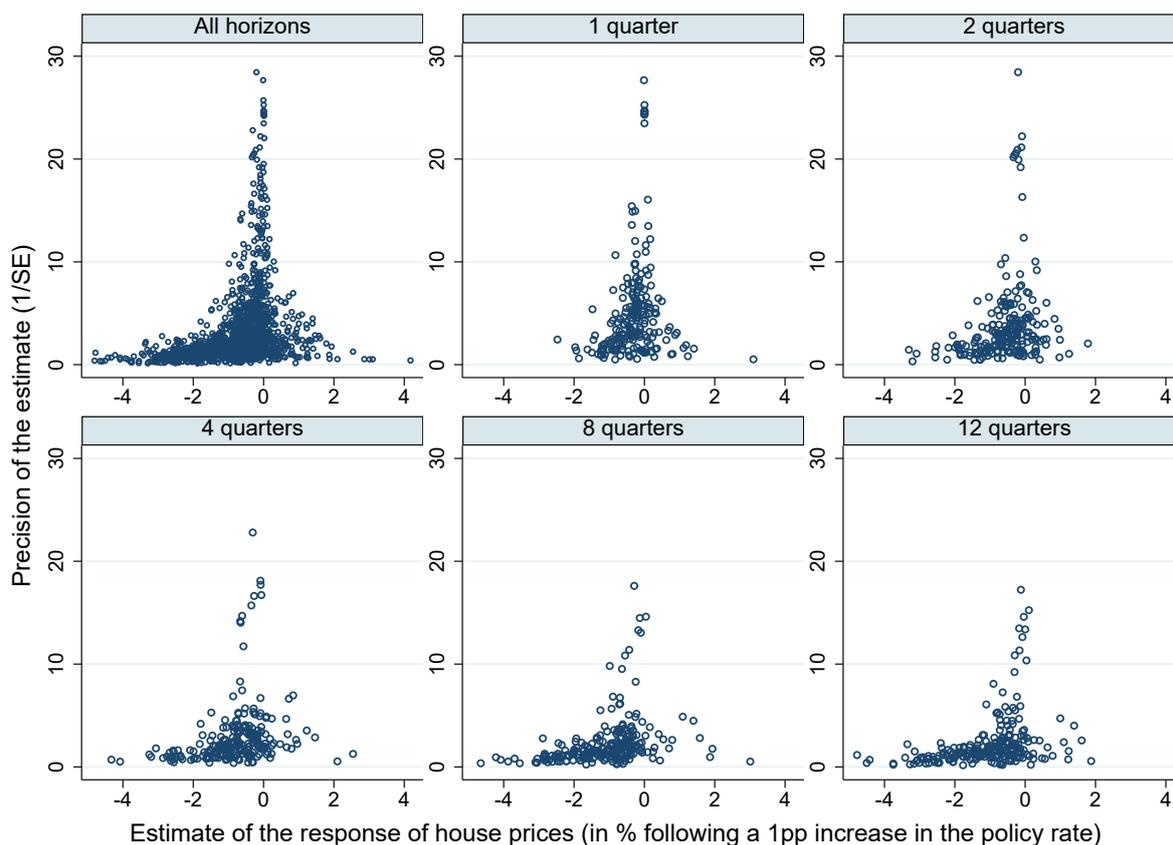
Publication bias is the systematic difference between the distribution of results produced by researchers and the distribution of results reported by researchers (both in working papers and journal articles). Sometimes the bias or its specific forms are also called selective reporting or p-hacking, though we prefer to work with the former, more general term. Whether or not publication bias is universally harmful is still a controversial question. On the one hand, it makes little sense to build a paper on unlikely results, such as those that suggest a rise in house prices following a monetary policy tightening.⁵ On the other hand, if such unlikely results are ignored, the literature as a whole gets biased upwards because it is hard to spot large estimates with the right sign and significance that are also due to chance or misspecifications. The resulting tension between the effects of publication selection at the micro and macro level is in the context of vector autoregressions nicely illustrated by the following quote due to Uhlig (2012, p. 38, emphasis added):

At a Carnegie-Rochester conference a few years back, Ben Bernanke presented an empirical paper, in which the conclusions nicely lined up with a priori reasoning about monetary policy. Christopher Sims then asked him, whether he would have presented the results, had they turned out to be at odds instead. His half-joking reply was, that he presumably would not have been invited if that had been so. There indeed is the *danger (or is it a valuable principle?)* that a priori economic theoretical biases filter the empirical evidence that can be brought to the table in the first place.

⁵Though such impulse response functions can still be explained, for example, by the theory of rational bubbles, see Galí (2014) and Galí & Gambetti (2015).

In experimental research, publication bias can in principle be tamed by preregistration (Olken, 2015; Strømland, 2019), and the American Economic Association has established a registry for experimental papers explicitly to “counter publication bias” (Siegfried, 2012, p. 648). Such registries are also common in medical research, where publication bias has long been recognized as a grave problem (Nosek *et al.*, 2018), but we are not aware of a field in which publication bias would be extirpated by preregistration. Perhaps publication bias is allowed to survive in many fields because at the micro level of individual studies it can really represent a valuable principle, a specification check that clearly tells the researcher that something is wrong with the model or the data. It is then the task for those who evaluate the literature as a whole to correct for the macro publication bias. As we have noted in the Introduction, our basic identification procedure is based on the Lombard effect. If estimation is imprecise and data are noisy, the researcher will need to try harder to produce estimates that are fully consistent with the intuition and theory—that is, statistically significant negative responses of house prices to a monetary tightening. So we expect more precise estimates to be less biased.

Figure 3: Funnel plots suggest publication bias



Notes: In the absence of publication bias the scatter plots should resemble inverted funnels symmetric around the most precise estimates. Outliers are omitted from the graphs for ease of exposition but included in all statistical tests.

The logic of the identification assumption can be described in a so-called funnel plot often used in medical research. The funnel plot is a scatter plot of estimate size (on the horizontal

axis) and estimate precision (on the vertical axis). The most precise estimates will be close to the underlying mean effect, while less precise estimates will be more dispersed, together forming the shape of an inverted funnel. If the mean underlying effect is not zero, the most precise estimates will always be statistically significant and therefore reported. In the absence of publication bias all imprecise estimates will be reported with the same probability. If publication bias is present and the literature as a whole prefers significant negative responses of house prices to a monetary tightening, then given the same precision positive (and small negative) estimates will be reported with a lower probability than large negative estimates, because the latter are more likely to be statistically significant. The funnel plots reported in Figure 3 show signs of asymmetry consistent with publication bias. It is interesting to observe that the degree of asymmetry increases as the horizon of the impulse response increases, perhaps reflecting the fact that insignificant estimates are less acceptable at longer horizons.

The asymmetry of the funnel can be tested explicitly (Card & Krueger, 1995; Egger *et al.*, 1997):

$$\hat{x}_{i,j} = \alpha_0 + \beta SE_{i,j} + \epsilon_{i,j}, \quad (2)$$

where $\hat{x}_{i,j}$ denotes the i -th estimated effect of interest rates on house prices in the j -th study, and SE is the corresponding standard error. Parameter α_0 denotes the mean effect beyond bias (that is, conditional on infinite precision and thus no publication selection), while β represents the intensity of publication bias. The simple regression has at least two problems (aside from ignoring heterogeneity, which we will address in the next section). First, it assumes a linear relationship between the standard error and the extent of publication bias. But the correlation between bias and precision can vary for different values of precision. When the estimate is very precise, small changes in precision do not alter the intensity of publication selection because they do not alter the designation of statistical significance at standard levels. When the estimate is very imprecise, small increases in precision do not achieve statistical significance and thus do not influence publication probability and selective reporting. It is for intermediate values of precision, and especially around the main threshold for statistical significance, that a relation between estimates and standard errors is more likely.

Second, (2) assumes that the standard error is exogenous. The assumption can be realistic in medical research where the standard error is basically given to the researcher (it is computed based on a straightforward formula of the number of observations), but in economics the computation of the standard error is a complex exercise. In any case the standard error is not given but can be influenced by the estimation approach; therefore publication bias can work through both point estimates and standard errors. A related problem is that the standard error itself is estimated, and thus (2) suffers from attenuation bias (Stanley, 2005).

We relax the linearity assumption by employing the stem-based method by Furukawa (2021) and the selection model by Andrews & Kasy (2019). The stem-based method (alluding to the stem of the funnel plot, a portion that includes precise estimates) is a nonparametric approach that optimizes the trade-off between bias and variance. When only the most precise studies are used to compute the mean effect, little publication bias remains, but the variance of the

Table 1: Linear and nonlinear tests suggest publication bias

Time after a monetary policy shock:	1 quarter	2 quarters	4 quarters	8 quarters	12 quarters	16 quarters
<i>PANEL A: Linear models</i>						
<i>Regression of reported estimates on their standard errors, ordinary least squares</i>						
Standard error (publication bias)	-0.815 [*] (0.463)	-1.117 ^{***} (0.367)	-1.353 ^{***} (0.427)	-1.160 ^{***} (0.308)	-0.667 ^{**} (0.317)	-0.375 [*] (0.215)
Constant (corrected mean effect)	[-1.220, -.015] -0.014 (0.168) [-.213, .145]	[-1.987, -.351] -0.020 (0.185) [-.417, .479]	[-2.373, -.410] -0.015 (0.248) [-.576, .739]	[-2.051, -.150] -0.233 (0.219) [-.790, .444]	[-1.542, .009] -0.501 ^{**} (0.252) [-1.072, .154]	[-1.281, .163] -0.560 ^{***} (0.201) [-.968, -.039]
<i>Regression of reported estimates on their standard errors, weighted by inverse variance</i>						
Standard error (publication bias)	-0.705 ^{***} (0.179)	-0.874 ^{***} (0.147)	-1.092 ^{***} (0.203)	-1.160 ^{***} (0.204)	-0.964 ^{***} (0.234)	-0.732 ^{***} (0.199)
Constant (corrected mean effect)	[-1.228, -.393] -0.059 (0.046) [-.058, .014]	[-1.207, -.586] -0.147 ^{**} (0.062) [-.303, .081]	[-1.555, -.730] -0.185 ^{**} (0.093) [-.360, .130]	[-1.629, -.710] -0.234 [*] (0.121) [-.496, .044]	[-1.534, -.497] -0.212 (0.135) [-.521, -.030]	[-1.343, -.337] -0.175 (0.116) [-.533, -.010]
<i>PANEL B: Nonlinear models</i>						
<i>Stem-based method (Furukawa, 2021)</i>						
Corrected mean effect	-0.017 (0.053)	-0.192 ^{***} (0.074)	-0.318 ^{***} (0.116)	-0.324 [*] (0.185)	-0.172 (0.145)	-0.146 ^{**} (0.071)
<i>Selection model (Andrews & Kasy, 2019)</i>						
Corrected mean effect, break at $t = 1.645$	-0.004 ^{***} (0.002)	-0.155 [*] (0.094)	-0.361 ^{***} (0.079)	-0.404 ^{**} (0.190)	-0.205 (0.187)	-0.065 (0.115)
Corrected mean effect, break at $t = 1$	-0.023 (0.027)	-0.008 (0.234)	-0.213 ^{**} (0.130)	-0.149 (0.194)	-0.182 ^{**} (0.110)	-0.030 (0.114)
<i>P-uniform* (van Aert & van Assen, 2021)</i>						
Corrected mean effect, break at $t = 1.645$	-0.133 ^{***}	-0.121 ^{***}	-0.140 ^{***}	-0.134 ^{***}	-0.118 ^{***}	-0.095 ^{***}
Corrected mean effect, break at $t = 1$	-0.072 ^{***}	-0.091 ^{***}	-0.103 ^{***}	-0.098 ^{***}	-0.093 ^{***}	-0.075 ^{***}
Observations	222	227	237	237	232	226

Notes: The mean uncorrected effect at the 8-quarter horizon was -1.2 . Standard errors, clustered at the level of studies and countries, are depicted in round brackets; confidence intervals from wild bootstrap are in square brackets. The p-uniform* method reports p-values, which are all below 0.001 and thus not shown in the table. The selection model and p-uniform* require specifying the break corresponding to a publication selection rule. The wild bootstrap (Cameron *et al.*, 2008) is implemented via the *boottest* package in Stata (Roodman *et al.*, 2019). *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

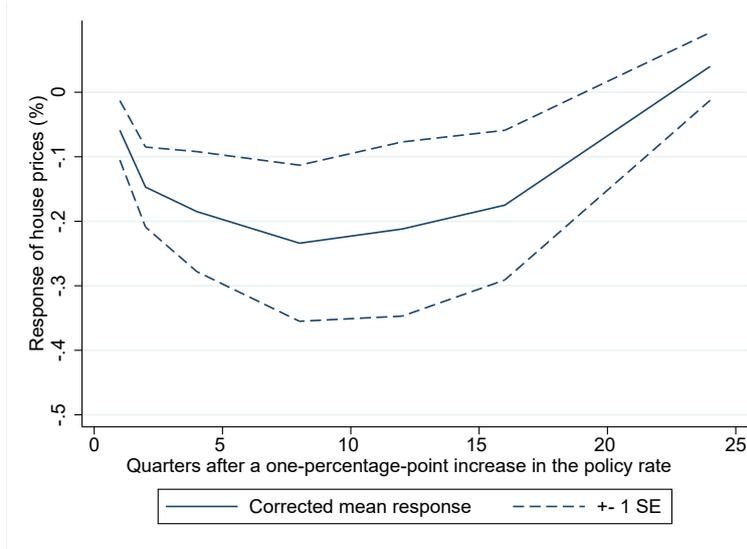
mean estimate increases because it is inefficient to discard information. When less precise studies are included as well, the variance of the mean estimate decreases, but the mean is more contaminated by bias. Furukawa (2021) presents a straightforward way how to weigh these two problems and select the optimal number of most precise studies for the computation of the mean effect. The method rests on minimizing the mean squared error, which can be expressed as the sum of publication bias (squared) and variance; the necessary assumption is that the most precise study does not suffer from any publication bias but is still subject to sampling error. The selection model by Andrews & Kasy (2019) assumes that the probability of reporting for each estimate depends on its sign and statistical significance, with changes in probability at 0 and the main thresholds for statistical significance. The model then re-weights the estimates based on the computed reporting probabilities.

We relax the exogeneity assumption by employing the p-uniform* method by van Aert & van Assen (2021). The method does not assume anything about the relationship between estimates and standard errors but uses the statistical principle that the distribution of p-values should be uniform at the underlying mean effect size. Consequently, it recomputes p-values and searches for the mean value of the semi-elasticity that would be consistent with a uniform distribution of p-values. In addition, in the online appendix we use several techniques that are robust to the exogeneity assumption but do not provide estimates of the mean semi-elasticity corrected for publication bias; instead they test for the presence of bias (Gerber & Malhotra, 2008a; Elliott *et al.*, 2022) or test the null hypothesis that the corrected effect is zero (Simonsohn *et al.*, 2014a).

The main results are shown in Table 1. Panel A reports the findings of linear models (the regression of estimates on standard errors), while Panel B focuses on nonlinear models. We employ double clustering of standard errors at the level of studies and countries. Because we only have 37 studies in our dataset, we additionally report confidence intervals based on wild bootstrap. In the first part of Panel A we run the regression specified in (2), while in the second part we run weighted least squares with weights proportional to the inverse variance of the reported estimates. The weighted specification corrects for the heteroskedasticity inherent in (2). In the case of the selection model and p-uniform* in Panel B we need to specify the relevant thresholds for statistical significance. As we have noted in the Introduction, it is common in the VAR literature to use the 68% confidence interval (that is, one standard error on both sides of the mean) instead of the 95% interval common elsewhere in economics. A few VAR studies use the 90% interval, so we set our thresholds for the corresponding values of the t-statistic at 1 and 1.645. Two observations emerge from the table. First, the corrected mean effect is always smaller than the simple mean. Second, the effect at the eight-quarter horizon is usually statistically significant even after correction for publication bias and ranges from -0.40 (selection model) to -0.10 (p-uniform*). The presence of publication bias and significance of the mean effect corrected for publication bias is further supported by robustness checks presented in the online appendix that use the techniques of Gerber & Malhotra (2008a), Simonsohn *et al.* (2014a), and Elliott *et al.* (2022).

The weighted least squares specification yields estimates of the mean effect close to the

Figure 4: Mean impulse response after correction for publication bias



Notes: Based on the weighted least squares specification reported in Panel A of Table 1.

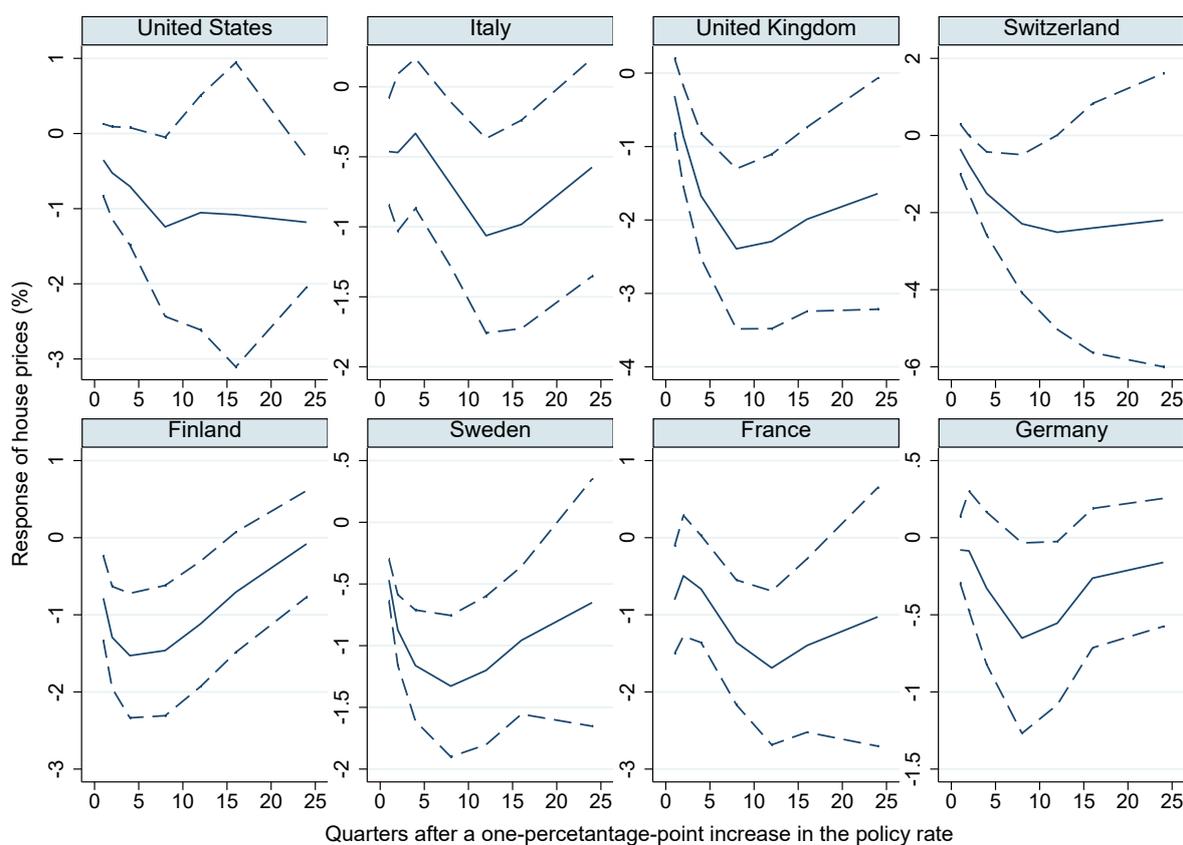
median of those of all the techniques considered, and we use this specification to construct the implied impulse response corrected for publication bias. The response is shown in Figure 4 and presents a similar shape to the one discussed earlier in relation to Figure 1: house prices decrease swiftly following a monetary policy tightening, the effect peaks after two years and then dissipates. The main difference is the size of the response, which is now much smaller: -0.23% after two years compared to the simple uncorrected mean estimate of -1.2% . Publication bias thus has important quantitative implications for the estimated effectiveness of monetary policy in taming house prices. Nevertheless, the finding of publication bias, and the mean impulse response itself, may be contaminated by the differences in the context in which the estimates are obtained. In the next section we thus turn to the heterogeneity in the estimates.

4 Heterogeneity

The previous literature has hinted on the differences in the transmission of monetary policy to house prices depending on the context of countries, time periods, and estimation techniques (among others, Iacoviello & Minetti, 2003; Assenmacher-Wesche & Gerlach, 2010; Bjørnland & Jacobsen, 2010; Calza *et al.*, 2013). But these studies could compare only a few countries, a few business cycles, and a few models computed using different specifications. Based on the efforts of these researchers, we build a large database of not only the reported results but also the factors that might have influenced those results. We are thus able to examine the heterogeneity in the response of house prices to policy rate shocks with much more power than the individual studies in the literature.

Consider Figure 5, which shows mean impulse responses reported for selected countries. While all the responses are intuitive and none shows the price puzzle, an increase in prices

Figure 5: Cross-country heterogeneity in transmission



Notes: The graphs show mean impulse responses reported for individual countries (the solid line) and 68% confidence intervals constructed by adding one standard error to each side of the mean (the dashed lines).

following a monetary policy tightening, the strength and speed of transmission varies greatly across countries. The maximum decrease in house prices following a one-percentage-point increase in the policy rate is -0.6% in Germany but -2.2% in the United Kingdom. In Finland house prices near their maximum response already after 2 quarters and dissipate quickly after two years, while the responses are persistent in Switzerland and the United States. The responses are quite precisely estimated for Finland and Germany, while transmission is uncertain in France, Switzerland, and the United States. In this section we try to explain these and other differences, together with evaluating the robustness of publication bias results to controlling for heterogeneity. Aside from variables that measure the characteristics of countries and the business cycle (what we call structural heterogeneity), we also control for the characteristics of data, specification, estimation, and publication. The definitions and summary statistics for all the variables are available in the online appendix; the variables are also briefly summarized below. For simplicity we focus on the four-quarter horizon, which is arguably the most relevant for monetary policy if the central bank intends to defuse a housing bubble in time (the results for the eight-quarter horizon are nevertheless similar).

4.1 Variables

Data Characteristics. We control for the characteristics of the data used in the primary studies. First, regarding data frequency, only around 11% of the estimates come from studies that use monthly data; the rest are based on quarterly data. Second, we control for whether simple time series (78% of all observations) or panel data are used in vector autoregressions. We are also interested in whether the strength of transmission changes over time, and we thus include the mean year of the dataset used. By doing so, we control for the potential change in transmission not accounted for by variables capturing structural heterogeneity, which will be described below. We also test whether the length of the sample used in the primary studies systematically affects the estimates.

Specification characteristics. When assessing the effect of monetary policy on the overall price level, Rusnak *et al.* (2013) find that study design has a significant effect on the results. For instance, they find that including output gap as a measure of output or commodity prices besides overall prices systematically affects the results. In a similar way, we create dummies for additional endogenous variables included in VAR models estimating the transmission of monetary policy to house prices. We include a dummy equal to one if the GDP deflator is used instead of the usual consumer price index. Next, we include dummy variables that equal one if a measure of credit (usually real credit to the private sector or mortgage loans) is used (28% of cases), if the long-term interest rate is used (20% of cases), and if consumption, residential investment, the money supply, the exchange rate, and the foreign interest rate are included. We distinguish between nominal and real house prices, though nominal house prices are used in merely around 6% of the studies. We only include studies which use residential house prices, not commercial house prices, land prices, or rent prices. As far as the remaining aspects of the estimation specification are concerned, we control for the number of lags included in the VAR model. The number of lags affects the persistence of the impulse responses and can thus also affect the strength of transmission.

Estimation characteristics. Another important dimension in which estimates differ is the estimation technique. The primary studies typically use a reduced-form VAR employing ordinary least squares or maximum likelihood, and they usually rely on recursive ordering as their identification scheme (76% of all estimates). We control for the use of sign restrictions. Since sign restrictions differ across papers (the restriction may not be imposed on all variables in the same direction), we distinguish between two cases that are important for the transmission to house prices. First, we include a dummy variable equal to one if sign restrictions are imposed on the house price variable, guaranteeing the expected sign. Second, we include a dummy if sign restrictions are imposed on any other variables, but not house prices. We then control for other types of nonrecursive identification (such as long-run restrictions) and, regarding the estimation procedure, we also create a dummy variable that equals one if a Bayesian VAR is estimated (around 15% of the estimates).

Publication characteristics. While the variables introduced above can help us control for some aspects of study quality, other aspects will remain difficult to code or even observe. As

additional proxies for quality, we include three publication characteristics. First, we control for the number of Google Scholar citations each study has received on average over the first three years after it appeared on Google Scholar for the first time. This way we take into account the long and variable publication lags in economics, where working papers might accumulate a significant amount of citations even prior to publication, and our chosen measure also accounts for the fact that per-year citations tend to decline as the paper gets older. We also include variables reflecting publication in a peer-reviewed journal and the RePEc discounted recursive impact factor of the outlet. We expect highly-cited studies published in peer-reviewed journals with a high impact factor to be of higher quality than other studies, *ceteris paribus*. A qualification is of course in order, because any potential correlation between the size of the estimates and the publication characteristics can be also due to publication bias and not necessarily due to genuine systematic effects of (unobserved) study quality on results. One must therefore be cautious with the interpretation of the results related to this group of variables.

Structural heterogeneity. We include a wide range of external variables (marked with the prefix “Country-level”), that is, variables obtained outside the primary studies to cover relevant macroeconomic, financial, demographic, and housing supply factors. For each impulse response, we compute these variables as mean values of the time span used to deliver the particular impulse response for a given country or a group of countries (in which case we weight the individual country-level values by country GDP). First, we include a measure of economic development—disposable income per capita. We also include variables capturing boom and crisis periods. Second, we include interest rate variables, which we suspect may interact with the transmission to house prices. We control for the level of the short-term interest rate itself: transmission can be more complete at higher (“normal”) monetary policy rates, while it can change at low interest rates because of excessive risk-taking by economic agents. On the other hand, very low interest rates or prolonged periods of very low interest rates may cause asymmetries in the transmission. In consequence, a prolonged period of low interest rates fueling credit and house price booms could be mirrored by a stronger reaction of house prices to monetary policy. Long-term interest rates (10-year government bond yields) are more relevant than the short-term rates for the transmission to house prices, and they are often driven by factors independent of the policy rate, such as demographics, inequality, savings glut, the relative price of capital, and amount of public investment (Rachel & Smith, 2017). Due to collinearity concerns, we include the term premium (spread) instead the long-term rate *per se*.⁶ We also include the inflation rate in the country: as shown by Rusnak *et al.* (2013), periods of high inflation are often associated with a lower credibility of the central bank and thus weaker transmission.

Third, we control for the characteristics of the lending market by including the credit-to-GDP ratio in order to account for the level of indebtedness as well as for the level of financial development. The inclusion of the mortgage-to-GDP ratio yields similar results, but because the amount of mortgages is unavailable for several countries in our dataset, we use the credit-to-GDP ratio instead to increase the number of degrees of freedom available for our analysis.

⁶Though, as noted by Gertler & Karadi (2015), the term premium itself is likely to be influenced by short-term rate changes.

We also include a variable capturing the share of mortgage loans with floating interest rates: the higher the share of floating-rate mortgages, the stronger the immediate transmission to the overall mortgage interest rate, and possibly the stronger the transmission to house prices in general. For similar reasons we also control for the average maturity of mortgage loans in the country. Fourth, regarding demographic characteristics we account for population growth in the country. If population growth is high, transmission may be weaker as house prices are driven by demographics rather than being affected by monetary policy.

Fifth, we include several characteristics of the housing sector. In order to account for house supply factors, we include the number of building permits. A low number of building permits indicates restricted housing supply and potentially hampered transmission of monetary policy.⁷ We also cover the home ownership structure. We include a proxy for tourism as a demand factor rather than a housing supply one. The remaining variables capturing structural heterogeneity relate to house prices themselves. In particular, we include the standardized price-to-income ratio as a proxy for overvaluation of house prices. The price-to-income ratio, available from the OECD database, is measured as the nominal house price divided by nominal disposable income per capita and can be considered a measure of affordability. As another potential proxy to capture overvalued house prices we include a variable capturing the number of periods house price growth is above its long-term average.

4.2 Estimation

We intend to find out whether the variables introduced above are systematically related to the reported effects of monetary policy on house prices. The easiest way would be to regress the estimated semi-elasticities on all the variables. But because of the large number of variables (39), such an estimation would be inefficient because many of the variables will probably not belong in the best underlying model. In other words, we face substantial model uncertainty, which is coupled with collinearity. Both can be addressed by Bayesian model averaging (BMA) with a dilution prior. BMA runs many models with different combinations of the explanatory variables and then constructs a weighted average over these models with weights proportional to model fit and complexity. The dilution prior (George, 2010) gives each model an additional weight proportional to the determinant of the correlation matrix, so that collinearity is penalized in the final output of Bayesian model averaging. BMA was pioneered in the social sciences by Raftery (1995) and Raftery *et al.* (1997) and recently used in meta-analysis, for example, by Havranek *et al.* (2018a,b,c), Bajzik *et al.* (2020), Cazachevici *et al.* (2020), Havranek & Sokolova (2020), Zigrainova *et al.* (2021), Gechert *et al.* (2022), and Matousek *et al.* (2022). As a robustness check, we use a frequentist alternative (frequentist model averaging, FMA), which is based on Magnus *et al.* (2010) and Amini & Parmeter (2012).

⁷The number of building permits acts as a proxy for housing supply. Other variables could serve this purpose: for example the number of dwellings. Nevertheless, the inclusion of this variable led to collinearity in our dataset. Another candidate variable would be an estimate of the sensitivity of house prices to housing supply. However, we are restricted by availability for our wide cross-country sample. Therefore, we stick to the number of building permits as a house supply proxy often used in the literature (e.g., Grimes & Aitken, 2010; Paciorek, 2013).

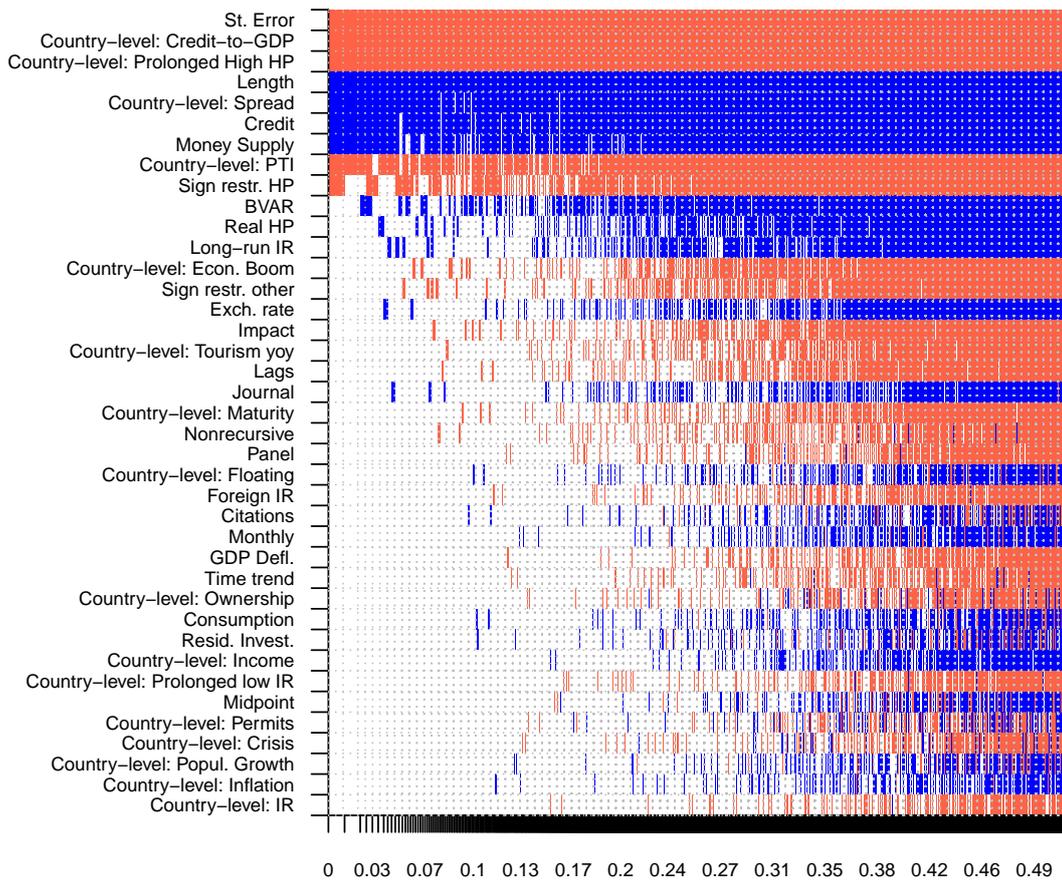
BMA can potentially run 2^{39} regressions with all the possible combinations of variables. Such a computation would take several months, and we avoid it by using the Markov Chain Monte Carlo process and its Metropolis-Hastings algorithm (Zeugner & Feldkircher, 2015), which goes through the most probable models. The posterior model probability then expresses the weight of each model. The estimated coefficients for every variable are weighted by the posterior model probability through all the models. For each variable we thus obtain a posterior inclusion probability (PIP), which denotes the sum of the posterior model probabilities of all the models in which the variable is included.

Concerning priors, in the baseline specification the unit information g-prior (UIP) recommended by Eicher *et al.* (2011) gives the prior the same weight as one observation of the data. It constitutes our benchmark setting, addressing the lack of prior knowledge regarding the parameter values. Moreover, the dilution prior addressing collinearity provides us with the benchmark model prior. Aside from the weight proportional to the determinant of the correlation matrix, all models have the same prior probability. As a robustness check of our baseline BMA results, we estimate BMA using alternative g-priors and model priors. We use a combination of the unit information g-prior and the uniform model prior and a combination of the Hannan-Quinn (HQ) g-prior and the random model prior (Fernandez *et al.*, 2001; Ley & Steel, 2009). As we have noted, we also use frequentist model averaging as an additional robustness check. In FMA we use Mallows's criterion for model averaging (Hansen, 2007), and the covariate space is orthogonalized using the approach of Amini & Parmeter (2012).

4.3 Results

Figure 6 summarizes the results of Bayesian model averaging graphically. Columns denote individual regression models from the best ones on the left, and the variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities; only the 10,000 best models are shown, which is why the cumulative probability does not run to 1. To ensure convergence we employ 3 million iterations and 1 million burn-ins. Blue color (darker in grayscale) means that the variable is included and the estimated sign is positive, i.e. transmission is weaker. Red color (lighter in grayscale) means that the variable is included and the estimated sign is negative, i.e. transmission is stronger. Blank cells denote exclusion of the variable. Eight variables are included in most of the best models, which means that these variables are effective in explaining the heterogeneity in the reported semi-elasticities: the standard error (a proxy for publication bias), credit to GDP (a proxy for financial development), prolonged growth in house prices and price-to-income ratio (proxies for the build-up of a housing bubble), length of the time series (a proxy for small-sample bias), the term premium (spread between short- and long-term rates, a proxy for risk-taking and position in the business cycle), and the inclusion of credit and money supply in the VAR (proxies for omitted variables, or alternatively variables introducing a masking problem because the effect of short-term rates on house prices may work through effects on liquidity). The remaining

Figure 6: Model inclusion in Bayesian model averaging



Notes: The response variable is the estimated effect of a one-percentage-point change in the interest rate on house prices after four quarters. Columns denote individual models; the variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities; only the 10,000 best models are shown, and they cover about 50% of posterior model probability. To ensure convergence we employ 3 million iterations and 1 million burn-ins. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive, i.e. transmission is weaker. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative, i.e. transmission is stronger. No color = the variable is not included in the model. The numerical results of the BMA exercise are reported in Table 2. A detailed description of the variables is available in the online appendix.

Table 2: Why reported impulse responses vary

Category	Variable	PIP	Post. mean	Post. SD
<i>Publication bias</i>	SE	1.000	-1.540	0.141
<i>Data characteristics</i>	Monthly	0.031	0.011	0.086
	Panel	0.042	-0.006	0.045
	Length	0.982	1.661	0.456
	Midpoint	0.026	0.005	0.054
<i>Specification characteristics</i>	GDP Defl.	0.031	-0.005	0.052
	Foreign IR	0.033	-0.006	0.077
	Credit	0.868	0.415	0.218
	Consumption	0.030	0.002	0.029
	Resid. Invest.	0.029	0.003	0.043
	Money Supply	0.683	0.433	0.342
	Exch. rate	0.113	0.028	0.096
	Long-run IR	0.163	0.065	0.173
	Real HP	0.173	0.088	0.224
	Lags	0.073	-0.004	0.020
	Time trend	0.030	-0.003	0.032
<i>Estimation characteristics</i>	BVAR	0.427	0.267	0.347
	Sign restr. HP	0.491	-0.387	0.453
	Sign restr. other	0.124	-0.067	0.219
	Nonrecursive	0.052	-0.011	0.066
<i>Publication characteristics</i>	Citations	0.032	0.001	0.015
	Impact	0.095	-0.016	0.063
	Journal	0.068	0.014	0.065
<i>Structural heterogeneity</i>	Country-level: Crisis	0.024	0.000	0.004
	Country-level: IR	0.014	-0.001	0.011
	Country-level: Prolonged low IR	0.027	-0.001	0.005
	Country-level: Spread	0.901	0.444	0.211
	Country-level: Floating	0.039	0.000	0.001
	Country-level: Tourism	0.080	-0.001	0.006
	Country-level: Income	0.028	0.013	0.117
	Country-level: Inflation	0.020	0.001	0.010
	Country-level: Credit-to-GDP	0.998	-0.011	0.003
	Country-level: Popul. Growth	0.022	0.002	0.037
	Country-level: PTI	0.648	-0.015	0.013
	Country-level: Prolonged High HP	0.983	-0.108	0.028
	Country-level: Permits	0.024	0.000	0.001
	Country-level: Maturity	0.063	-0.022	0.115
	Country-level: Ownership	0.030	0.000	0.002
	Country-level: Econ. Boom	0.154	-0.007	0.019
Observations	Constant 225	1.000	-1.553	NA

Notes: PIP = posterior inclusion probability. SD = standard deviation. Variables with a posterior inclusion probability higher than 0.5 are shown in bold. We employ the unit information g-prior as recommended by Eicher *et al.* (2011) and the dilution prior to address collinearity (George, 2010). A detailed description of the variables is available in the online appendix.

variables have posterior inclusion probabilities below 0.5, which means they are not important in explaining the differences in reported results.

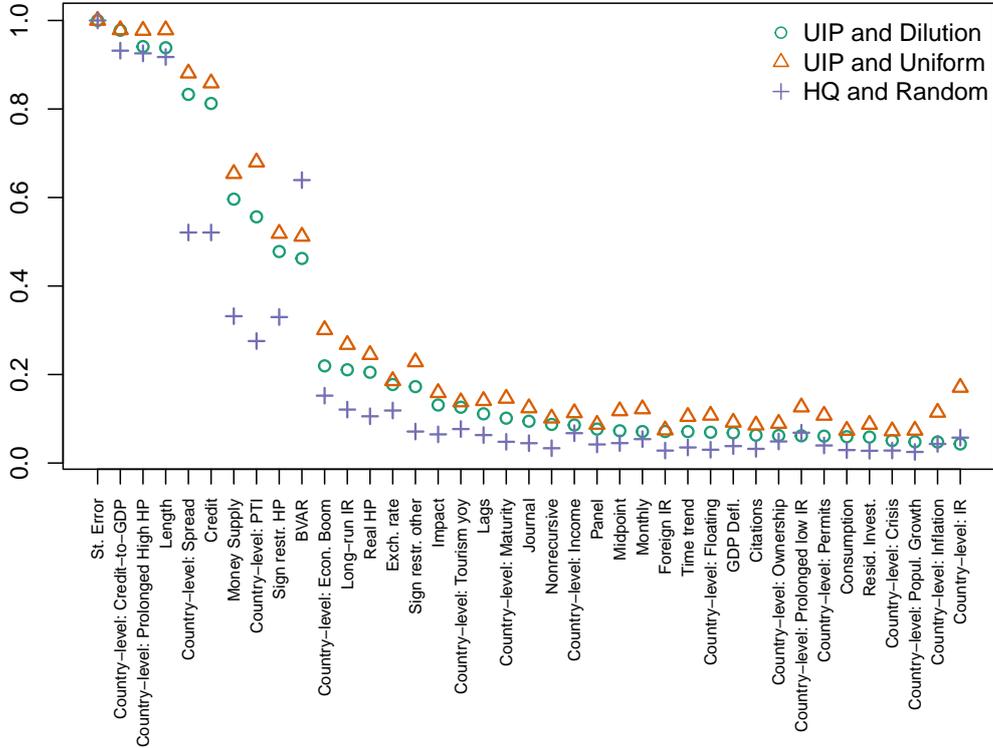
The numerical results of Bayesian model averaging are reported in Table 2. The eight variables with posterior inclusion probabilities above 0.5 are shown in bold. The posterior means presented in the table measure the partial derivatives of the reported semi-elasticities with respect to the variables in question. Our results suggest that the finding of substantial publication bias is robust to controlling for heterogeneity. Not only that the variable proves to be important in BMA, but it also has the largest posterior inclusion probability and the estimated coefficient (posterior mean) is larger than that reported in the previous section. We conclude that our previous finding of publication bias was not driven by omitting factors associated with heterogeneity. Next, we find that studies using longer time series are likely to report evidence of weaker transmission from monetary policy decisions to house prices. The result is consistent with a small-sample bias towards more negative semi-elasticities.

We find that specification characteristics are important for the reported estimates of the semi-elasticity. When credit or money supply are omitted from the analysis, the reported response of house prices tends to be more negative. There are two possible interpretations of the result. First, studies that exclude credit or money supply may suffer from omitted-variable bias, because they ignore the effect that liquidity itself has on house prices. Second, studies that include credit or money supply may suffer from the masking problem, because changes in the short-run rate affect the liquidity variables, which may then appear in a VAR to affect house prices on their own, masking the true causal effect of the policy rate. We also find that studies which put a (negative) sign restriction on the response of house prices tend to find, on average, more negative effects (even though the PIP here is slightly below 0.5). That finding is intuitive because with sign restrictions the price puzzle is a priori impossible. Note that sign restrictions on the response of house prices are used only by 5% of the specifications in our sample, and the results (including those on publication bias) would not be affected by excluding these restricted estimates entirely from our analysis.

Finally, our results suggest that variables reflecting the country's financial system and position in the cycle are important in explaining the reported semi-elasticities. The credit-to-GDP ratio has a PIP of almost 1 and shows a negative correlation with the reported response of house prices. Note that the result would be similar if we used the mortgage-to-GDP ratio instead. We opt for the former because data on the amount of mortgages are not available for every country and period of our dataset, so using mortgages would mean throwing away data. We interpret the finding, in line with Calza *et al.* (2013), as evidence for stronger transmission in countries with more developed mortgage (and, in general, credit) markets.

Next, we find that a flatter yield curve and an ongoing build-up of a bubble in the housing market are both associated with stronger transmission. The result is consistent with monetary policy being more effective at influencing house prices at the latter part of the business cycle, when banks and households are more prone to excess optimism and risk-taking, and adds some credence to the policy of leaning against the wind. In the next subsection, however, we show that

Figure 7: Sensitivity to alternative priors



Notes: UIP = unit information prior; the prior has the same weight as one observation of data. Uniform model prior = each model has the same prior weight. Dilution model prior = the prior weight of each model is proportional to the determinant of the correlation matrix. The HQ prior asymptotically mimics the Hannan-Quinn criterion. The random model prior assign the same prior weight to each model size (e.g., models with 10 variables have the same prior probability as models with 11 variables). PIP = posterior inclusion probability.

even under the best of circumstances the strength of transmission is insufficient to substantially mitigate housing bubbles. Moreover, as documented by Schularick *et al.* (2021), interest rate hikes during a house price boom can often lead to a financial crisis. A crisis precipitates a decrease in house prices, but the overall outcome is of course not what the policy maker had in mind. Unfortunately our dataset does not allow us to fully disentangle VAR evidence on the aforementioned mechanism from soft landings. We partially address this issue by including an interaction of the *Crisis* and *Prolonged high HP* variables (Figure C2 in the online appendix). The interaction is not important in BMA, which is in line with the interpretation that financial crises are not what drives our result that monetary policy becomes more effective at taming house prices after they have increased for several years.

As we have noted, we run several robustness checks to test the robustness of our results. Figure 7 shows the posterior inclusion probabilities for individual variables using different sets of priors in Bayesian model averaging. The changes are small and would not change our conclusions. In the online appendix we show the results of frequentist model averaging, Bayesian model

averaging for all semi-elasticities (not just those at the four-quarter horizon), and ordinary least squares regressions for all horizons separately. The results of FMA are broadly consistent with those of BMA, though generally yield less significance (for example, the p-values associated with the variables reflecting the inclusion of credit is 0.2). On the other hand, BMA and OLS results for all semi-elasticities imply more significance for most variables compared to our baseline BMA for semi-elasticities at the four-quarter horizon. In all cases, the finding of publication bias is statistically significant at the 1% level (in frequentist techniques) or has a posterior inclusion probability of 1 (in Bayesian techniques).

4.4 Implied Response

As the bottom line of our analysis we compute the impulse response implied by the entire literature but conditional on the absence of publication bias and potential misspecifications. We construct both the mean impulse response for the typical country and also responses for individual countries. In general, our results can be used to derive an implied impulse response conditional on any selected aspect of the financial system, business cycle, and estimation techniques. Technically the implied responses are computed as fitted values using the results of Bayesian model averaging and a definition of the preferred values for each variable included in BMA (or the sample mean if no preference can be made). So we plug in zero for the standard error in order to condition the implied response on the correction for publication bias. While we have noted that the linear correction for publication bias using the exogeneity assumption for the standard error is problematic in theory, we have also shown in the previous section that in the literature on monetary transmission to house prices the linear correction gives results similar to more complex methods. Since it is implausible to use the more complex methods of publication bias correction in BMA, we rely on the linear regression.

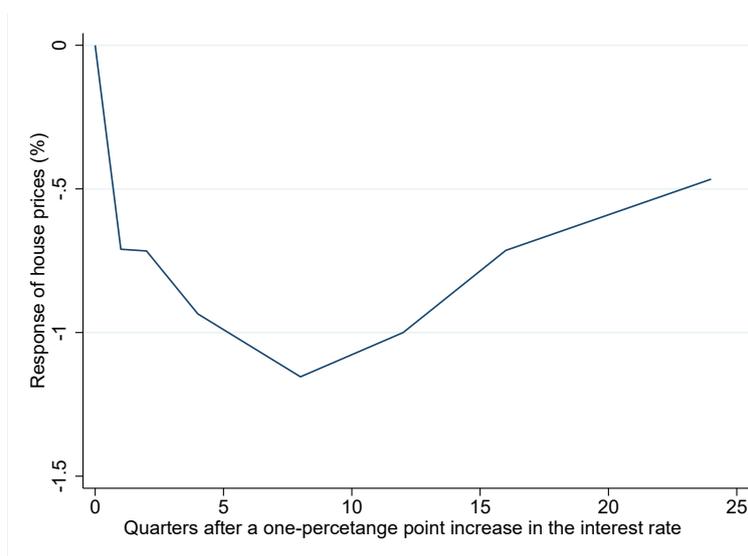
Table 3: Implied semi-elasticities

Horizon of the corresponding impulse response:	1Q	2Q	4Q	8Q	12Q	16Q
Agnostic on control variables (baseline)	-0.710	-0.716	-0.935	-1.154	-1.000	-0.714
Control for credit, money supply, and long-run IR	-0.016	-0.023	-0.241	-0.461	-0.306	-0.021
Excluding credit, money supply, and long-run IR	-0.920	-0.927	-1.145	-1.365	-1.210	-0.925
Finland	-0.576	-0.583	-0.801	-1.021	-0.866	-0.581
France	-1.477	-1.484	-1.702	-1.922	-1.767	-1.482
Germany	-0.497	-0.503	-0.722	-0.941	-0.787	-0.501
Italy	0.093	0.087	-0.132	-0.352	-0.197	0.088
United Kingdom	-1.344	-1.351	-1.569	-1.789	-1.634	-1.349
United States	-0.910	-0.916	-1.135	-1.354	-1.200	-0.914

Notes: The values represent the percentage response of house prices to a one-percentage-point increase in the policy rate. They correspond to mean estimates conditional on selected characteristics of the individual studies (see text for more details) and are computed based on fitted values from Bayesian model averaging (for example, by substituting “0” for the standard error, sample maximum for the number of citations, and so on). The estimates for individual countries are based on the baseline definition in the first row.

In order to put more weight on studies that use recent data we employ sample maximum for the variable capturing the mean year of data. Regarding specification characteristics, we prefer if the study uses real house prices (instead of nominal). Considering the controls for the long-term interest rate, credit, and money supply, we have no preference: as we have noted, the inclusion of these variables may create a masking problem, but their exclusion may lead to omitted-variable bias. Regarding estimation characteristics, we prefer Bayesian techniques and nonrecursive identification (structural VAR or sign restrictions). Regarding publication characteristics, we prefer highly cited studies published in peer-reviewed journals with a high impact factor: so we plug in 1 for the dummy variable reflecting journal publication and sample maxima for the number of per-year citations and the RePEc discounted recursive impact factor of the journal. We leave all other variables, including variables capturing structural heterogeneity, at sample means—of course, in the case of impulse responses constructed for individual countries we set the structural variables to the values corresponding to the individual countries.

Figure 8: Implied impulse response



Notes: The figure shows the mean impulse response reported in the literature and conditional on preferred aspects of data, methods, and publication. Based on the baseline exercise computed in Table 3.

The results are shown in Table 3 and Figure 8. While the main analysis in this section is based on the four-quarter horizon for ease of exposition, in order to compute the implied impulse response we need to run BMA analyses for each horizon separately. The corresponding analyses are not reported here, but in Table C3 in the online appendix we present the concise results of OLS estimates for each horizon. For each horizon the implied semi-elasticity is computed using the approach described in the previous two paragraphs. Note that confidence intervals or other traditional measures of statistical significance cannot be constructed, because the implied estimates are derived from Bayesian model averaging. The first row of the table shows the baseline implied response, which is also depicted graphically in Figure 8. The mean maximum corrected

semi-elasticity is -1.2 , which suggests, in practical terms, insufficiently strong transmission of monetary policy to house prices—on average at least.

The second row of Table 3 presents the results of the same exercise with the exception of the preferred values for specification characteristics. When one gives more weight to the omitted-variables interpretation and believes that liquidity and long-term interest rates should be controlled for in the VAR, the resulting semi-elasticity is much smaller in magnitude (-0.5 after two years). Next, we focus on the masking interpretation, according to which credit, money supply, and long-term rates should not be included in the VAR. The resulting responses of house prices are substantially larger with the semi-elasticity reaching -1.4 after eight quarters. Still the response is not large enough to be of practical importance in taming housing bubbles. It follows that different specification of the VAR model can easily change the estimated response of house prices by around one percentage point. In the remaining rows of Table 3 we compute impulse responses for several selected countries. Even the strongest semi-elasticity (-1.9 in France) is insufficient for plausible leaning against the wind when house prices inflation reaches double digits. The weakest semi-elasticity appears again in Italy (-0.4 after two years), which means that country-level characteristics can explain differences of up to 1.5 in the semi-elasticity. The differences explained by country and business-cycle characteristics can rise up to 3 if we select extreme values for these characteristics (not reported in the table). But even the impulse responses implied by the most extreme outliers in the values of these characteristics suggest semi-elasticities above -3 .

5 Concluding Remarks

We collect 1,555 estimates of the reaction of house prices to a monetary policy shock at different horizons reported in 237 impulse responses from 37 studies. Our results suggest that a one-percentage-point increase in the policy rate is on average associated with a maximum decrease of 1.2% in house prices after two years. We find that transmission varies substantially across countries and time: it is stronger in countries with more developed mortgage markets and in the latter part of the business cycle. But even the most optimistic estimates for the periods and countries with characteristics conducive to more effective transmission imply semi-elasticities of less than 3 in absolute value. So while leaning against the wind may help partly mitigate housing bubbles, the policy rate is a crude instrument for such a task and one costly in terms of inflation and unemployment. Svensson (2017) compares the benefits and costs of leaning against the wind and comes to the conclusion that in most contexts costs outweigh benefits by a large margin. Targeted macroprudential policy tools in the form of binding loan-to-income or debt-service-to-income ratios appear more likely to succeed in steering house prices, although empirical evidence on their effectiveness is still relatively thin (Poghosyan, 2020).

Three qualifications of our results are in order. First, in a way unusual but not unheard of in meta-analysis (Fabo *et al.*, 2021), we collect data from graphical results (impulse responses and the corresponding confidence intervals). Even though we do our best to codify the numerical values as precisely as possible, a random classical measurement error inevitably arises. In a

regression of the estimated semi-elasticity on the corresponding standard error, therefore, the slope coefficient is biased downward due to attenuation bias. Because in our benchmark models the slope coefficient measures the strength of publication bias, many of our estimations are likely to underestimate the effects of the bias and hence produce conservative corrections. In fact, however, the problem with measurement error is more benign in the synthesis of graphical results than in the traditional synthesis of numerical results. The reason is that numerical results are rounded. Because different studies round differently, measurement error might not be random across studies. Bruns *et al.* (2019) show that rounding can create a false impression of publication bias (for example, the clustering of t-statistics at integers such as 2).

Second, the baseline meta-analysis models that we use come from or are inspired by medical research. In medical research, it is common to assume that the standard error is given to the researcher, often directly proportional to the number of subjects. That is, the standard error is exogenous and in the absence of publication bias there should be no correlation between estimates and standard errors. But in economics the computation of the standard error forms an important part of the exercise: in the VAR literature, for example, the confidence intervals can be constructed using different bootstrapping approaches, and different estimation techniques will generally yield different intervals. It follows that publication bias can also work via unintentional manipulation of the reported precision, not only the reported point estimate as is commonly assumed in meta-analysis. One solution is to use a function of the number of observations as an instrument for the standard error, but in the VAR literature the instrument is weak. We thus employ the new p-uniform* technique (van Aert & van Assen, 2021) developed in psychology, which uses the distribution of p-values and assumes nothing about the relationship between estimates and standard errors. As robustness checks we also use the techniques by Gerber & Malhotra (2008a), Simonsohn *et al.* (2014a), and Elliott *et al.* (2022) that too do not need the exogeneity assumption.

Third, in this meta-analysis we ignore the growing literature on the effects of unconventional monetary policy on house prices (see, for example, Rahal, 2016; Lenza & Slacalek, 2018; Rosenberg, 2019). While the short-term policy rate appears to have only limited influence on house prices, other tools of monetary policy (such as quantitative easing) might have played a more prominent role recently. Indeed, our results indicate that controlling for liquidity reduces the reported effects of policy rates on house prices, which suggests that tools which primarily affect liquidity can be important. But the studies focusing on unconventional policy are quantitatively incomparable with the rest of our sample, and we believe they are best analyzed separately in a future research synthesis. The literature also lacks a thorough synthesis on the effects of macroprudential policies on house prices. As the body of relevant empirical research grows, conducting a meta-analysis will soon be possible in both realms.

References

- VAN AERT, R. C. & M. VAN ASSEN (2021): “Correcting for publication bias in a meta-analysis with the p-uniform* method.” *Working paper*, Tilburg University.
- AMINI, S. M. & C. F. PARMETER (2012): “Comparison of model averaging techniques: Assessing growth determinants.” *Journal of Applied Econometrics* **27(5)**: pp. 870–876.
- ANDREWS, I. & M. KASY (2019): “Identification of and correction for publication bias.” *American Economic Review* **109(8)**: pp. 2766–2794.
- ASPACHS-BRACONS, O. & M. P. RABANAL (2011): “The effects of housing prices and monetary policy in a currency union.” *IMF WP/11/6*, International Monetary Fund.
- ASSENMACHER-WESCHE, K. & S. GERLACH (2008a): “Financial structure and the impact of monetary policy on asset prices.” *SNB Working Papers 2008-16*, Swiss National Bank.
- ASSENMACHER-WESCHE, K. & S. GERLACH (2008b): “Monetary policy, asset prices and macroeconomic conditions: A panel-VAR study.” *NBB Working Paper No. 149*, National Bank of Belgium.
- ASSENMACHER-WESCHE, K. & S. GERLACH (2009): “Financial structure and the impact of monetary policy on property prices.” *Unpublished manuscript*, Goethe University, Frankfurt.
- ASSENMACHER-WESCHE, K. & S. GERLACH (2010): “Monetary policy and financial imbalances: Facts and fiction.” *Economic Policy* **25(63)**: pp. 437–482.
- ASTAKHOV, A., T. HAVRANEK, & J. NOVAK (2019): “Firm size and stock returns: A quantitative survey.” *Journal of Economic Surveys* **33(5)**: pp. 1463–1492.
- BAJZIK, J., T. HAVRANEK, Z. IRSOVA, & J. SCHWARZ (2020): “Estimating the Armington elasticity: The importance of data choice and publication bias.” *Journal of International Economics* **127(C)**: p. 103383.
- BAUER, G. & E. GRANZIERA (2017): “Monetary policy, private debt, and financial stability risks.” *International Journal of Central Banking* **13(3)**: pp. 337–373.
- BELKE, A., W. ORTH, & R. SETZER (2008): “Sowing the seeds for the subprime crisis: Does global liquidity matter for housing and other asset prices?” *International Economics and Economic Policy* **5(4)**: pp. 403–424.
- BENATI, L. (2021): “Leaning against house prices: A structural VAR investigation.” *Journal of Monetary Economics* **118(C)**: pp. 399–412.
- BERLEMANN, M. & J. FREESE (2013): “Monetary policy and real estate prices: A disaggregated analysis for Switzerland.” *International Economics and Economic Policy* **10(4)**: pp. 469–490.
- BJØRNLAND, H. C. & D. H. JACOBSEN (2010): “The role of house prices in the monetary policy transmission mechanism in small open economies.” *Journal of Financial Stability* **6(4)**: pp. 218–229.
- BLANCO-PEREZ, C. & A. BRODEUR (2020): “Publication bias and editorial statement on negative findings.” *Economic Journal* **130(629)**: pp. 1226–1247.
- BRODEUR, A., N. COOK, & A. HEYES (2020): “Methods matter: P-hacking and publication bias in causal analysis in economics.” *American Economic Review* **110(11)**: pp. 3634–60.
- BRODEUR, A., M. LÉ, M. SANGNIER, & Y. ZYLBERBERG (2016): “Star wars: The empirics strike back.” *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.
- BROWN, A. L., T. IMAI, F. VIEIDER, & C. CAMERER (2022): “Meta-analysis of empirical estimates of loss-aversion.” *Journal of Economic Literature* (**forthcoming**).
- BRUNS, S. B., I. ASANOV, R. BODE, M. DUNGER, C. FUNK, S. M. HASSAN, J. HAUSCHILDT, D. HEINISCH, K. KEMPA, J. KÖNIG *et al.* (2019): “Reporting errors and biases in published empirical findings: Evidence from innovation research.” *Research Policy* **48(9)**: p. 103796.
- BRUNS, S. B. & J. P. A. IOANNIDIS (2016): “P-curve and p-hacking in observational research.” *PloS ONE* **11(2)**: p. e0149144.
- BULLIGAN, G. (2010): “Housing and the macroeconomy: The Italian case.” In O. DDE BANDT, T. KNETSCH, J. PEÑALOSA, & F. ZOLLINO (editors), “Housing Markets in Europe: A Macroeconomic Perspective,” pp. 19–38. Springer.
- CALZA, A., T. MONACELLI, & L. STRACCA (2007): “Mortgage markets, collateral constraints, and monetary policy: Do institutional factors matter?” *CFS Working Paper No. 2007/10*, Center for Financial Studies, Goethe University Frankfurt.
- CALZA, A., T. MONACELLI, & L. STRACCA (2013): “Housing finance and monetary policy.” *Journal of the European Economic Association* **11(1)**: pp. 101–122.
- CAMERON, A. C., J. B. GELBACH, & D. L. MILLER (2008): “Bootstrap-based improvements for inference with clustered errors.” *The Review of Economics and Statistics* **90(3)**: pp. 414–427.
- CARD, D., J. KLUVE, & A. WEBER (2018): “What works? a meta analysis of recent active labor market program evaluations.” *Journal of the European Economic Association* **16(3)**: pp. 894–931.
- CARD, D. & A. B. KRUEGER (1995): “Time-series minimum-wage studies: A meta-analysis.” *The American Economic Review* **85(2)**: pp. 238–243.
- CARSTENSEN, K., O. HÜLSEWIG, &

- T. WOLLMERSHÄUSER (2009): “Monetary policy transmission and house prices: European cross country evidence.” *Cesifo working paper no. 2750*, Center for Economic Studies and ifo Institute (CESifo), University of Munich.
- CATTANEO, M. D., M. JANSSON, & X. MA (2020): “Simple local polynomial density estimators.” *Journal of the American Statistical Association* **115**(531): pp. 1449–1455.
- CAZACHEVICI, A., T. HAVRANEK, & R. HORVATH (2020): “Remittances and economic growth: A meta-analysis.” *World Development* **134**: p. 105021.
- CHRISTENSEN, G. & E. MIGUEL (2018): “Transparency, reproducibility, and the credibility of economics research.” *Journal of Economic Literature* **56**(3): pp. 920–980.
- COIBION, O., Y. GORODNICHENKO, L. KUENG, & J. SILVIA (2017): “Innocent bystanders? Monetary policy and inequality.” *Journal of Monetary Economics* **88**: pp. 70–89.
- DEL NEGRO, M. & C. OTROK (2007): “99 luftballons: Monetary policy and the house price boom across US states.” *Journal of Monetary Economics* **54**(7): pp. 1962–1985.
- DELLAVIGNA, S. & E. LINOS (2022): “RCTs to scale: Comprehensive evidence from two nudge units.” *Econometrica* **90**(1): pp. 81–116.
- DELLAVIGNA, S., D. POPE, & E. VIVALTI (2019): “Predict science to improve science.” *Science* **366**(6464): pp. 428–429.
- DEMARY, M. (2010): “The interplay between output, inflation, interest rates and house prices: International evidence.” *Journal of Property Research* **27**(1): pp. 1–17.
- DIAS, D. A. & J. B. DUARTE (2019): “Monetary policy, housing rents, and inflation dynamics.” *Journal of Applied Econometrics* **34**(5): pp. 673–687.
- EGGER, M., G. D. SMITH, & C. MINDER (1997): “Bias in meta-analysis detected by a simple, graphical test.” *Journal of Economic Surveys* **315**(7109): p. 629–634.
- EICHER, T. S., C. PAPAGEORGIOU, & A. E. RAFTERY (2011): “Default priors and predictive performance in bayesian model averaging, with application to growth determinants.” *Journal of Applied Econometrics* **26**(1): pp. 30–55.
- ELBOURNE, A. (2008): “The UK housing market and the monetary policy transmission mechanism: An SVAR approach.” *Journal of Housing Economics* **17**(1): pp. 65–87.
- ELLIOTT, G., N. KUDRIN, & K. WUTHRICH (2022): “Detecting p-hacking.” *Econometrica* **90**(2): pp. 887–906.
- FABO, B., M. JANCOKOVA, E. KEMPF, & L. PASTOR (2021): “Fifty Shades of QE: Comparing Findings of Central Bankers and Academics.” *Journal of Monetary Economics* **120**: pp. 1–20.
- FERNANDEZ, C., E. LEY, & M. F. J. STEEL (2001): “Model uncertainty in cross-country growth regressions.” *Journal of Applied Econometrics* **16**(5): pp. 563–576.
- FRATANTONI, M. & S. SCHUH (2003): “Monetary policy, housing, and heterogeneous regional markets.” *Journal of Money, Credit and Banking* **35**(4): pp. 557–589.
- FURUKAWA, C. (2021): “Publication bias under aggregation frictions: Theory, evidence, and a new correction method.” *Working paper*, MIT.
- GALÍ, J. (2014): “Monetary policy and rational asset price bubbles.” *American Economic Review* **104**(3): pp. 721–752.
- GALÍ, J. & L. GAMBETTI (2015): “The effects of monetary policy on stock market bubbles: Some evidence.” *American Economic Journal: Macroeconomics* **7**(1): pp. 233–257.
- GECHERT, S., T. HAVRANEK, Z. IRSOVA, & D. KOLCUNOVA (2022): “Measuring capital-labor substitution: The importance of method choices and publication bias.” *Review of Economic Dynamics* **45**: pp. 55–82.
- GEORGE, E. I. (2010): “Dilution priors: Compensating for model space redundancy.” In “Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown,” pp. 158–165. Institute of Mathematical Statistics.
- GERBER, A. & N. MALHOTRA (2008a): “Do statistical reporting standards affect what is published? publication bias in two leading political science journals.” *Quarterly Journal of Political Science* **3**(3): pp. 313–326.
- GERBER, A. S. & N. MALHOTRA (2008b): “Publication bias in empirical sociological research: Do arbitrary significance levels distort published results?” *Sociological Methods & Research* **37**(1): pp. 3–30.
- GERTLER, M. & P. KARADI (2015): “Monetary policy surprises, credit costs, and economic activity.” *American Economic Journal: Macroeconomics* **7**(1): pp. 44–76.
- GIULIODORI, M. (2005): “Monetary policy shocks and the role of house prices across european countries.” *Scottish Journal of Political Economy* **52**(4): pp. 519–543.
- GRIMES, A. & A. AITKEN (2010): “Housing supply, land costs and price adjustment.” *Real Estate Economics* **38**(2): pp. 325–353.
- GUPTA, R., M. JURGILAS, A. KABUNDI, & S. M. MILLER (2012a): “Monetary policy and housing sector dynamics in a large-scale bayesian vector autoregressive model.” *International Journal of Strategic Property Management* **16**(1): pp. 1–20.

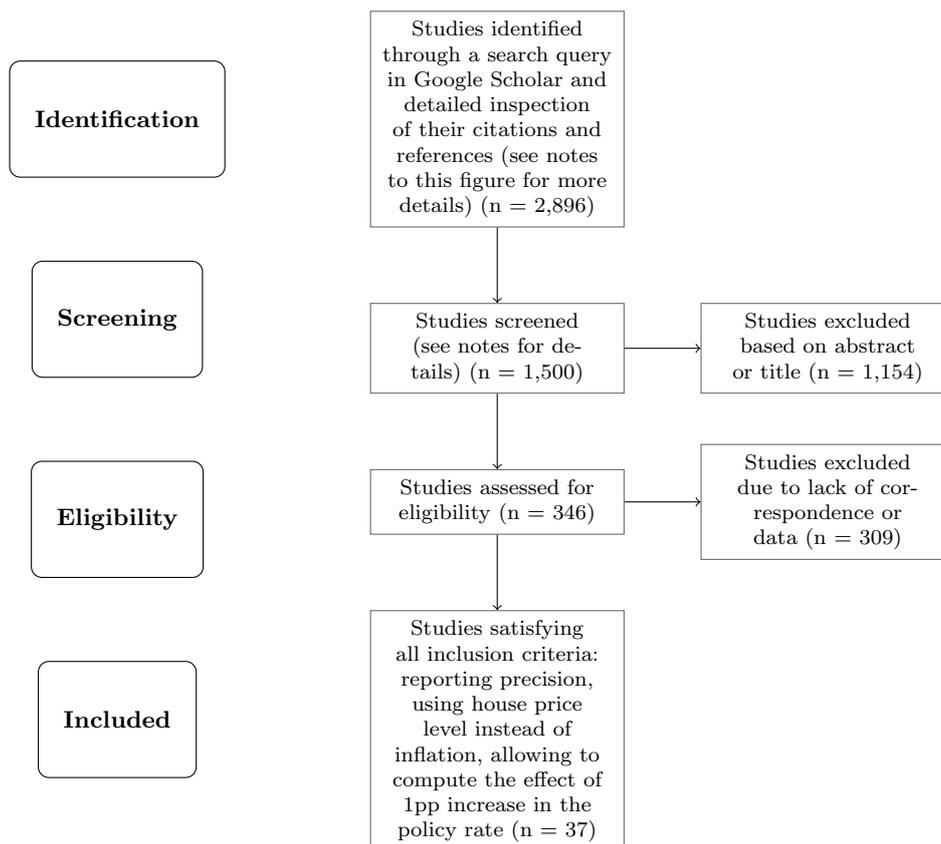
- GUPTA, R., M. JURGILAS, S. M. MILLER, & D. VAN WYK (2012b): "Financial market liberalization, monetary policy, and housing sector dynamics." *International Business & Economics Research Journal (IBER)* **11(1)**: pp. 69–82.
- HANSEN, B. E. (2007): "Least squares model averaging." *Econometrica* **75(4)**: p. 1175–1189.
- HAVRANEK, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.
- HAVRANEK, T., D. HERMAN, , & Z. IRSOVA (2018a): "Does daylight saving save electricity? a meta-analysis." *The Energy Journal* **39(2)**: pp. 35–61.
- HAVRANEK, T., Z. IRSOVA, & T. VLACH (2018b): "Measuring the income elasticity of water demand: The importance of publication and endogeneity biases." *Land Economics* **94(2)**: pp. 259–283.
- HAVRANEK, T., Z. IRSOVA, & O. ZEYNALOVA (2018c): "Tuition fees and university enrolment: A meta-regression analysis." *Oxford Bulletin of Economics and Statistics* **80(6)**: pp. 1145–1184.
- HAVRANEK, T. & A. SOKOLOVA (2020): "Do consumers really follow a rule of thumb? three thousand estimates from 144 studies say "probably not"." *Review of Economic Dynamics* **35**: pp. 97–122.
- HAVRANEK, T., T. D. STANLEY, H. DOUCOULIAGOS, P. BOM, J. GEYER-KLINGEBERG, I. IWASAKI, W. R. REED, K. ROST, & R. C. M. VAN AERT (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* **34(3)**: pp. 469–475.
- IACOVIELLO, M. (2002): "House prices and business cycles in europe: A VAR analysis." *Bc working papers in economics 540*, Boston College.
- IACOVIELLO, M. & R. MINETTI (2003): "Financial liberalization and the sensitivity of house prices to monetary policy: Theory and evidence." *The Manchester School* **71(1)**: pp. 20–34.
- IACOVIELLO, M. & R. MINETTI (2008): "The credit channel of monetary policy: Evidence from the housing market." *Journal of Macroeconomics* **30(1)**: pp. 69–96.
- IACOVIELLO, M. & S. NERI (2010): "Housing market spillovers: Evidence from an estimated DSGE model." *American Economic Journal: Macroeconomics* **2(2)**: pp. 125–164.
- IMAI, T., T. A. RUTTER, & C. F. CAMERER (2021): "Meta-analysis of present-bias estimation using convex time budgets." *Economic Journal* **131(636)**: pp. 1788–1814.
- IOANNIDIS, J. P., T. STANLEY, & H. DOUCOULIAGOS (2017): "The power of bias in economics research." *Economic Journal* **127(605)**: pp. 236–265.
- IWASAKI, I. (2022): "The finance-growth nexus in Latin America and the Caribbean: A meta-analytic perspective." *World Development* **149(C)**.
- JANNSEN, N., G. POTJAGAILO, & M. H. WOLTERS (2015): "Monetary policy during financial crises: Is the transmission mechanism impaired?" *International Journal of Central Banking* **15(4)**.
- JAROCINSKI, M. & F. SMETS (2008): "House prices and the stance of monetary policy." *Federal Reserve Bank of St. Louis Review* **90(Jul)**: pp. 339–366.
- LENZA, M. & J. SLACALEK (2018): "How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area." *Working Paper Series 2190*, European Central Bank.
- LEY, E. & M. F. STEEL (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression." *Applied Econometrics* **24**: pp. 651–674.
- MAGNUS, J. R., O. POWELL, & P. PRUFER (2010): "A comparison of two model averaging techniques with an application to growth empirics." *Journal of Econometrics* **154(2)**: pp. 139–153.
- MATOUSEK, J., T. HAVRANEK, & Z. IRSOVA (2022): "Individual discount rates: a meta-analysis of experimental evidence." *Experimental Economics* **25(1)**: pp. 318–358.
- MCCLOSKEY, D. & S. T. ZILIAK (2019): "What quantitative methods should we teach to graduate students? a comment on Swann's "Is precise econometrics an illusion?"." *The Journal of Economic Education* **50(4)**: pp. 356–361.
- MCDONALD, J. F. & H. H. STOKES (2013): "Monetary policy, mortgage rates and the housing bubble." *Economics & Finance Research* **1(1)**: pp. 82–91.
- MIRANDA-AGRIPPINO, S. & H. REY (2020): "US monetary policy and the global financial cycle." *The Review of Economic Studies* **87(6)**: pp. 2754–2776.
- MUSSO, A., S. NERI, & L. STRACCA (2011): "Housing, consumption and monetary policy: How different are the US and the euro area?" *Journal of Banking & Finance* **35(11)**: pp. 3019–3041.
- NCUBE, M. & E. NDOU (2011): "Monetary policy transmission, house prices and consumer spending in South Africa: An SVAR approach." *Working paper 133*, African Development Bank Group.
- NEISSER, C. (2021): "The elasticity of taxable income: A meta-regression analysis." *Economic Journal* **131(640)**: pp. 3365–3391.
- NOSEK, B. A., C. R. EBERSOLE, A. C. DEHAVEN, & D. T. MELLOR (2018): "The preregistration revolution." *Proceedings of the National Academy of Sciences* **115(11)**: pp. 2600–2606.
- OLKEN, B. A. (2015): "Promises and perils of pre-analysis plans." *Journal of Economic Perspectives* **29(3)**: pp. 61–80.
- PACIOREK, A. (2013): "Supply constraints and housing market dynamics." *Journal of Urban Economics* **77**:

- pp. 11–26.
- POGHOSYAN, T. (2020): “How effective is macroprudential policy? Evidence from lending restriction measures in EU countries.” *Journal of Housing Economics* **49(C)**.
- POWELL, T. & D. WESSEL (2021): “Why is the New Zealand government telling its central bank to focus on rising house prices?” Brookings, Hutchins Center on Fiscal and Monetary Policy, April 2, 2021.
- RACHEL, L. & T. D. SMITH (2017): “Are low real interest rates here to stay?” *International Journal of Central Banking* **13(3)**: pp. 1–42.
- RAFTERY, A. E. (1995): “Bayesian model selection in social research.” *Sociological Methodology* pp. 111–163.
- RAFTERY, A. E., D. MADIGAN, & J. A. HOETING (1997): “Bayesian model averaging for linear regression models.” *Journal of the American Statistical Association* **92(437)**: pp. 179–191.
- RAHAL, C. (2016): “Housing markets and unconventional monetary policy.” *Journal of Housing Economics* **32**: pp. 67–80.
- ROODMAN, D., M. Ø. NIELSEN, J. G. MACKINNON, & M. D. WEBB (2019): “Fast and wild: Bootstrap inference in stata using boottest.” *The Stata Journal* **19(1)**: pp. 4–60.
- ROSENBERG, S. (2019): “The effects of conventional and unconventional monetary policy on house prices in the Scandinavian countries.” *Journal of Housing Economics* **46(C)**.
- RUSNAK, M., T. HAVRANEK, & R. HORVATH (2013): “How to solve the price puzzle? a meta-analysis.” *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.
- SÁ, F., P. TOWBIN, & T. WIELADEK (2011): “Low interest rates and housing booms: The role of capital inflows, monetary policy and financial innovation.” *FRB Working Paper No. 79*, Federal Reserve Bank of Dallas.
- SÁ, F. & T. WIELADEK (2011): “Monetary policy, capital inflows, and the housing boom.” *Cambridge working papers in economics 1141*, University of Cambridge.
- SCHULARICK, M., L. TER STEEGE, & F. WARD (2021): “Leaning against the wind and crisis risk.” *American Economic Review: Insights* **3(2)**: pp. 199–214.
- SIEGFRIED, J. J. (2012): “Minutes of the Meeting of the Executive Committee: Chicago, IL, January 5, 2012.” *American Economic Review* **102(3)**: pp. 645–652.
- SIMONSOHN, U., L. D. NELSON, & J. P. SIMMONS (2014a): “P-curve: A key to the file-drawer.” *Journal of Experimental Psychology: General* **143(2)**: pp. 534–547.
- SIMONSOHN, U., L. D. NELSON, & J. P. SIMMONS (2014b): “P-curve and effect size: Correcting for publication bias using only significant results.” *Perspectives on Psychological Science* **9(6)**: pp. 666–681.
- SIMS, C. A. (1980): “Macroeconomics and reality.” *Econometrica* **48(1)**: pp. 1–48.
- SOUSA, R. M. (2014): “Wealth, asset portfolio, money demand and policy rule.” *Bulletin of Economic Research* **66(1)**: pp. 95–111.
- STANLEY, T. D. (2001): “Wheat from chaff: Meta-analysis as quantitative literature review.” *Journal of Economic Perspectives* **15(3)**: pp. 131–150.
- STANLEY, T. D. (2005): “Beyond publication bias.” *Journal of Economic Surveys* **19(3)**: pp. 309–345.
- STANLEY, T. D. (2008): “Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection.” *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.
- STANLEY, T. D. & H. DOUCOULIAGOS (2010): “Picture this: A simple graph that reveals much ado about research.” *Journal of Economic Surveys* **24(1)**: pp. 170–191.
- STANLEY, T. D. & H. DOUCOULIAGOS (2017): “Neither fixed nor random: Weighted least squares meta-regression.” *Research Synthesis Methods* **8(1)**: pp. 19–42.
- STANLEY, T. D., H. DOUCOULIAGOS, & J. P. IOANNIDIS (2017): “Finding the power to reduce publication bias.” *Statistics in medicine* **36(10)**: pp. 1580–1598.
- STEEL, M. F. (2020): “Model averaging and its use in economics.” *Journal of Economic Literature* **58(3)**: pp. 644–719.
- STRÖMLAND, E. (2019): “Preregistration and reproducibility.” *Journal of Economic Psychology* **75(PA)**.
- SVENSSON, L. E. (2014): “Inflation Targeting and ‘Leaning against the Wind’.” *International Journal of Central Banking* **10(2)**: pp. 103–114.
- SVENSSON, L. E. (2017): “Cost-benefit analysis of leaning against the wind.” *Journal of Monetary Economics* **90(C)**: pp. 193–213.
- TAYLOR, J. B. (2007): “Housing and monetary policy.” *NBER Working Paper No. 13682*, National Bureau of Economic Research.
- UGUR, M., S. A. CHURCHILL, & H. M. LUONG (2020): “What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power.” *Research Policy* **49(1)**.
- UHLIG, H. (2005): “What are the effects of monetary policy on output? Results from an agnostic identification procedure.” *Journal of Monetary Economics* **52(2)**: pp. 381–419.
- UHLIG, H. (2012): “Economics and reality.” *Journal of Macroeconomics* **34(1)**: pp. 29–41.

- VARGAS-SILVA, C. (2008): “Monetary policy and the US housing market: A VAR analysis imposing sign restrictions.” *Journal of Macroeconomics* **30(3)**: pp. 977–990.
- WADUD, I. M., O. H. BASHAR, & H. J. A. AHMED (2012): “Monetary policy and the housing market in Australia.” *Journal of Policy Modeling* **34(6)**: pp. 849–863.
- WILLIAMS, J. C. (2016): “Measuring the effects of monetary policy on house prices and the economy.” *BIS Papers No. 88*, Bank for International Settlements.
- WU, L. & Y. BIAN (2018): “Housing, consumption and monetary policy: How different are the first-, second- and third-tier cities in China?” *Applied Economics Letters* **25(15)**: pp. 1107–1111.
- XUE, X., W. R. REED, & A. MENCLOVA (2020): “Social capital and health: a meta-analysis.” *Journal of Health Economics* **72(C)**.
- ZEUGNER, S. & M. FELDKIRCHER (2015): “Bayesian model averaging employing fixed and flexible priors: The BMS package for R.” *Journal of Statistical Software* **68(4)**: pp. 1–37.
- ZIGRAIOVA, D., T. HAVRANEK, Z. IRSOVA, & J. NOVAK (2021): “How puzzling is the forward premium puzzle? A meta-analysis.” *European Economic Review* **134(C)**.

A Details of Literature Search

Figure A1: PRISMA flow diagram



Notes: Our baseline search query in Google Scholar is (“house” OR “housing” OR “real estate” OR “property” OR “residential”) AND “prices” AND (“monetary policy” OR “interest rate” OR “MP” OR “IR” OR “funds rate” OR “FF rate” OR “macroprudential” OR “financial stability”). We collect the first 500 studies returned by this search and call them the baseline set. In addition we collect 500 studies most often cited by the studies in the baseline set and 500 studies that most often cite the studies in the baseline set. We are left with 1,500 studies that we screen. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is an evidence-based set of items for reporting in systematic reviews and meta-analyses. More details on PRISMA and reporting standard of meta-analysis in general are provided by Havranek *et al.* (2020).

Table A1: Studies included in the dataset

Author	Title	Year
Iacoviello (2002)	House Prices and Business Cycles in Europe: a VAR Analysis	2002
Iacoviello & Minetti (2003)	Financial liberalization and the sensitivity of house prices to monetary policy: theory and evidence	2003
Giuliodori (2005)	Monetary Policy Shocks and the Role of House Prices Across European Countries	2004
Calza <i>et al.</i> (2007)	Mortgage markets, collateral constraints, and monetary policy: do institutional factors matter?	2007
Assenmacher-Wesche & Gerlach (2008a)	Financial structure and the impact of monetary policy on asset prices	2008
Assenmacher-Wesche & Gerlach (2008b)	Monetary policy, asset prices and macroeconomic conditions: a panel-VAR study	2008
Belke <i>et al.</i> (2008)	Sowing the seeds for the subprime crisis: does global liquidity matter for housing and other asset prices?	2008
Elbourne (2008)	The UK housing market and the monetary policy transmission mechanism: An SVAR approach	2008
Iacoviello & Minetti (2008)	The credit channel of monetary policy: Evidence from the housing market	2008
Jarocinski & Smets (2008)	House prices and the stance of monetary policy	2008
Vargas-Silva (2008)	Monetary policy and the US housing market: A VAR analysis imposing sign restrictions	2008
Assenmacher-Wesche & Gerlach (2009)	Financial structure and the impact of monetary policy on property prices	2009
Carstensen <i>et al.</i> (2009)	Monetary policy transmission and house prices: European cross-country evidence	2009
Assenmacher-Wesche & Gerlach (2010)	Monetary policy and financial imbalances: facts and fiction	2010
Bjørnland & Jacobsen (2010)	The role of house prices in the monetary policy transmission mechanism in small open economies	2010
Bulligan (2010)	Housing and the macroeconomy: the Italian case	2010
Demary (2010)	The interplay between output, inflation, interest rates and house prices: international evidence	2010
Aspachs-Bracons & Rabanal (2011)	The effects of housing prices and monetary policy in a currency union	2011
Musso <i>et al.</i> (2011)	Housing, consumption and monetary policy: how different are the US and the euro area?	2011
Ncube & Ndou (2011)	Monetary policy transmission, house prices and consumer spending in South Africa: An SVAR approach	2011
Sá <i>et al.</i> (2011)	Low interest rates and housing booms: the role of capital inflows, monetary policy and financial innovation	2011
Sá & Wieladek (2011)	Monetary policy, capital inflows and the housing boom	2011
Gupta <i>et al.</i> (2012a)	Monetary policy and housing sector dynamics in a large-scale Bayesian vector autoregressive model	2012
Gupta <i>et al.</i> (2012b)	Financial Market Liberalization, Monetary Policy, and Housing Sector Dynamics	2012
Wadud <i>et al.</i> (2012)	Monetary policy and the housing market in Australia	2012
Berleemann & Freese (2013)	Monetary policy and real estate prices: a disaggregated analysis for Switzerland	2013
Calza <i>et al.</i> (2013)	Housing finance and monetary policy	2013
McDonald & Stokes (2013)	Monetary policy, mortgage rates and the housing bubble	2013
Sousa (2014)	Wealth, asset portfolio, money demand and policy rule	2014
Bauer & Granziera (2017)	Monetary policy, private debt and financial stability risks	2017
Coibion <i>et al.</i> (2017)	Innocent bystanders? Monetary policy and inequality in the U.S.	2017
Wu & Bian (2018)	Housing, consumption and monetary policy: how different are the first-, second- and third-tier cities in China?	2018
Dias & Duarte (2019)	Monetary policy, housing rents, and inflation dynamics	2019
Jannsen <i>et al.</i> (2015)	Monetary policy during financial crises: Is the transmission mechanism impaired?	2019
Rosenberg (2019)	The effects of conventional and unconventional monetary policy on house prices in the Scandinavian countries	2019
Miranda-Agrippino & Rey (2020)	U.S. monetary policy and the global financial cycle	2020
Benati (2021)	Leaning against house prices: A structural VAR investigation	2021

Appendix B Extensions of Publication Bias Models (for Online Publication)

Appendix B.1 Caliper Test

As an extension to the previously reported tests of publication bias we apply the caliper test as proposed in Gerber & Malhotra (2008a) and Gerber & Malhotra (2008b) and recently implemented by Bruns *et al.* (2019). The caliper test is based on the analysis of discontinuities in the reported t-statistics: if no selective reporting is present, there should be no discontinuities around the conventional significance thresholds. In other words the number of t-statistics reported in the literature just above the threshold (“over caliper”) should not be statistically different from the number of reported t-statistics just below the threshold (“under caliper”). The test does not allow us to compute the true effect beyond bias but serves as an indicator of whether publication selection appears in the literature, thus providing us with a robustness check of the previous results. The results are presented in Table B1. Primarily we examine the significance threshold corresponding to the 68% confidence interval: although the threshold is usually much stricter in the empirical literature featuring point estimates, in the case of VAR models and impulse response functions the 68% confidence interval is the most frequently reported (almost 70% of our estimates use it), so we suspect that publication selection could be related to this threshold. We use caliper sizes of 0.1, 0.3, and 0.5. The results show that publication selection is present at the one-quarter horizon with a wider caliper. If we test the parameter against the value of 0.4 (i.e., a 60:40 distribution around the thresholds, instead of 50:50, as reasoned in Bruns *et al.* 2019), then evidence of publication selection is stronger and common for most horizons. This is broadly in line with our previous results on publication selection.

Table B1: Results of the caliper test

Caliper size	All horizons	1 quarter	2 quarters	4 quarters	8 quarters	12 quarters	16 quarters
0.1	<i>0.505</i>	<i>0.688</i>	0.500	0.444	0.429	0.467	0.474
(95% LCI)	[0.404]	[0.432]	[0.244]	[0.0393]	[-0.0658]	[0.181]	[0.226]
0.3	<i>0.520</i>	<i>0.604</i>	0.524	<i>0.553</i>	<i>0.625</i>	0.478	0.436
(95% LCI)	[0.467]	[0.461]	[0.366]	[0.406]	[0.494]	[0.328]	[0.301]
0.5	<i>0.516</i>	0.658	0.474	<i>0.551</i>	<i>0.598</i>	0.500	0.425
(95% LCI)	[0.475]	[0.549]	[0.359]	[0.438]	[0.493]	[0.385]	[0.319]

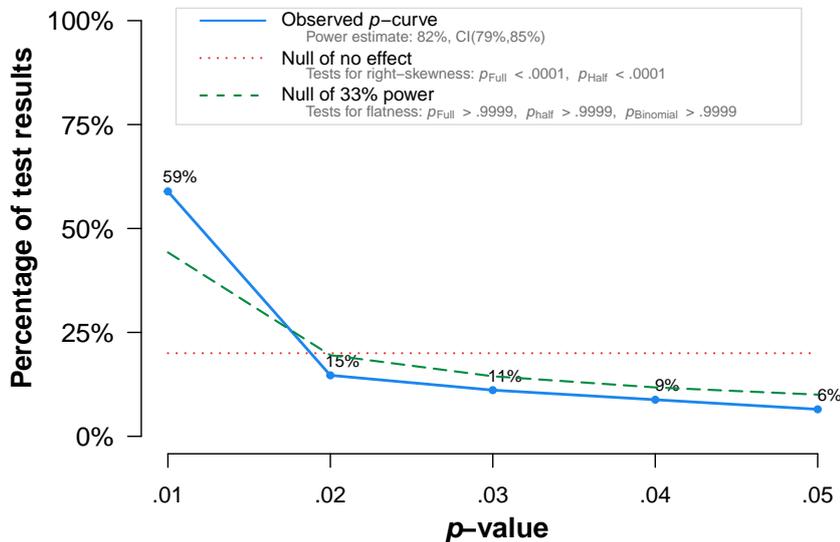
Notes: The table shows the results of the caliper test for three caliper sizes 0.1, 0.3, and 0.5. The reported numbers represent the share of observations in the narrow interval that are above the significance threshold. LCI = lower bound of the confidence interval. The test parameter is the following: $C = \frac{n_{oc}}{n_{oc} + n_{uc}}$, where n_{oc} and n_{uc} stand for the number of observations with t-statistics in the interval above the threshold (“over caliper”) and below the threshold (“under caliper”). For the significance threshold we use the criterion of one standard error above the estimate (commonly used in the VAR literature). The one-sided hypothesis $H_0 : C \leq 0.5$ is tested against $H_1 : C > 0.5$. 95% lower confidence intervals for the test parameters are reported in parenthesis. Significant caliper test results when testing $H_0 : C \leq 0.5$ are shown in bold; significant caliper test results when testing $H_0 : C \leq 0.4$ are shown in italics.

Appendix B.2 Tests Based on the Distribution of p-values

Appendix B.2.1 p-curve

Now we look at the distribution of p-values. First, we employ the p-curve method, which is primarily intended to test the null hypothesis that the literature has no evidential value (that is, no effect of monetary policy on house prices beyond publication bias). The technique was developed by Simonsohn *et al.* (2014a) and Simonsohn *et al.* (2014b). Based on Figure B1, we obtain evidence for evidential value, which is consistent with a right-skewed distribution, while a left-skewed distribution would suggest p-hacking. In addition to contrast to the common p-curve we also plot the whole distribution of p-values (not only those significant up to the 5% significance level) in Figure B2 to see whether there are distinct jumps at different thresholds associated with conventional statistical significance. There is no clear evidence for such jumps.

Figure B1: p-curve results

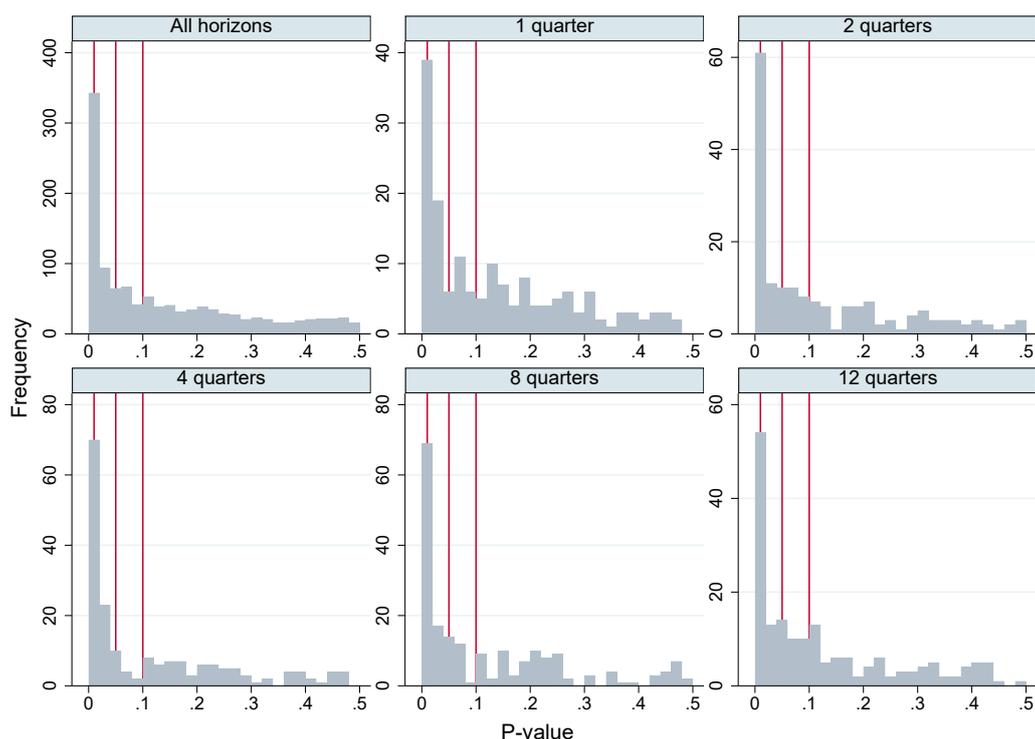


Note: The observed p-curve includes 477 statistically significant ($p < .05$) results, of which 379 are $p < .025$. There were 1078 additional results entered but excluded from p-curve because they were $p > .05$.

Appendix B.2.2 Tests introduced by Elliott *et al.* (2022)

Elliott *et al.* (2022) analyze p-hacking based on the distribution of p-values and introduce novel testable restrictions. They show that the p-curve (distribution of p-values across studies) is i) non-increasing and continuous in the absence of p-hacking, ii) completely monotone, with upper bounds on p-curve. In their empirical application they use binomial, Fisher's, and density discontinuity tests, as already used before in Simonsohn *et al.* (2014a) and Cattaneo *et al.* (2020). Besides that, they also develop new, more powerful tests: a histogram-based test for non-increasingness, a histogram-based test for 2-monotonicity and bounds, and least concave majorant (LCM) test based on concavity of the CDF of p-values. The results of these tests are available in Table B2. All of the tests have null hypothesis of no p-hacking. While with

Figure B2: Distribution of p-values



less powerful tests (binomial and Fisher) we do not reject the null of no p-hacking, we can reject it with the test for non-increasingness (CS1), 2-monotonicity (CS2B) and also density discontinuity test at horizons between 2 and 12 quarters in all cases. As in the main body of the paper, we run the tests at a threshold of $t=1$, instead of 1.96, as this is the most common threshold in impulse responses of VAR models.

Table B2: Tests used by Elliott *et al.* (2022)

Test	All	1 Quarter	2 Quarters	4 Quarters	8 Quarters	12 Quarters	16 Quarters
Binomial	0.990	0.702	0.500	0.837	1.000	0.820	0.500
Fisher	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Discontinuity	0.000	0.161	0.000	0.000	0.001	0.015	0.671
CS1	0.069	0.891	0.053	0.000	0.019	0.000	0.194
CS2B	0.000	0.000	0.008	0.000	0.000	0.000	0.004
LCM	1.000	0.987	0.999	0.996	0.691	0.997	0.994
N near $t=1$	178	32	21	26	35	19	23
N	1054	142	156	174	182	164	146

Notes: CS1 is the test for non-increasingness. CS2B is the test for K-monotonicity. LCM is the test based on the concavity of the CDF of p-values. Values in bold indicate rejections of the hypothesis of no p-hacking.

Table B3: Results concerning publication bias remain similar when studies on house price inflation are included

Time after a monetary policy shock:	1 quarter	2 quarters	4 quarters	8 quarters	12 quarters	16 quarters
<i>PANEL A: Linear models</i>						
<i>Regression of reported estimates on their standard errors, ordinary least squares</i>						
Standard error (publication bias)	-0.846** (0.397)	-1.082*** (0.369)	-1.319*** (0.432)	-1.152*** (0.309)	-0.675** (0.313)	-0.386* (0.211)
	[-1.702, 0.253]	[-1.981, -0.328]	[-2.382, -0.325]	[-2.070, -0.231]	[-1.561, -0.009]	[-1.298, 0.143]
Constant (corrected mean effect)	-0.038 (0.141)	-0.071 (0.185)	-0.081 (0.261)	-0.293 (0.246)	-0.550* (0.287)	-0.616** (0.246)
	[-0.392, 0.372]	[-0.464, 0.448]	[-0.663, 0.713]	[-0.939, 0.378]	[-1.186, 0.164]	[-1.142, 0.011]
<i>Regression of reported estimates on their standard errors, weighted by inverse variance</i>						
Standard error (publication bias)	-0.745*** (0.174)	-0.877*** (0.137)	-1.112*** (0.175)	-1.215*** (0.160)	-1.028*** (0.185)	-0.797*** (0.168)
	[-1.157, -0.328]	[-1.206, -0.569]	[-1.551, -0.712]	[-1.657, -0.811]	[-1.536, -0.484]	[-1.331, -0.269]
Constant (corrected mean effect)	-0.079* (0.047)	-0.176*** (0.058)	-0.212** (0.087)	-0.241** (0.108)	-0.215* (0.123)	-0.178 (0.120)
	[-0.217, 0.026]	[-0.352, 0.048]	[-0.464, 0.089]	[-0.572, 0.092]	[-0.575, 0.141]	[-0.524, 0.211]
<i>PANEL B: Nonlinear models</i>						
<i>Stem-based method (Furukawa, 2021)</i>						
Corrected mean effect	-0.012 (0.119)	-0.122 (0.191)	-0.168 (0.263)	-0.128 (0.331)	-0.187 (0.130)	-0.146** (0.075)
<i>Selection model (Andrews & Kasy, 2019)</i>						
Corrected mean effect, break at $t = 1.645$	-0.125*** (0.045)	-0.104 (0.319)	-0.340*** (0.065)	-0.412*** (0.096)	-0.208 (0.142)	-0.068 (0.097)
Corrected mean effect, break at $t = 1$	0.005 (0.070)	-0.050 (0.130)	-0.187*** (0.060)	-0.208 (0.163)	-0.196 (0.162)	-0.070 (0.131)
<i>P-uniform* (van Aert & van Assen, 2021)</i>						
Corrected mean effect, break at $t = 1.645$	-0.164***	-0.142***	-0.159***	-0.147***	-0.124***	-0.093***
Corrected mean effect, break at $t = 1$	-0.097***	-0.110***	-0.121***	-0.110***	-0.104***	-0.080***
Observations	240	245	255	255	249	241

Notes: The results are computed using the original dataset and, additionally, estimates from 8 studies on house prices inflation recomputed using the approach of Fabo *et al.* (2021). Standard errors, clustered at the level of studies and countries, are depicted in round brackets; confidence intervals from wild bootstrap are in square brackets. The p-uniform* method reports p-values, which are all below 0.001 and thus not shown in the table. The selection model and p-uniform* require specifying the break corresponding to a publication selection rule. The wild bootstrap (Cameron *et al.*, 2008) is implemented via the *boottest* package in Stata (Roodman *et al.*, 2019). *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Appendix C Summary Statistics and Extensions of Heterogeneity Models (for Online Publication)

Table C1: Description and summary statistics of regression variables

Label	Description	Mean	SD
Estimate	The reported effect of a one-percentage-point increase in the interest rate on house prices (after four quarters in %).	-0.853	1.242
Standard Error	The reported or implied standard error of the estimate.	0.780	0.894
<i>Data characteristics</i>			
Monthly	= 1 if the data were collected at the monthly frequency (reference category: quarterly data).	0.112	0.315
Panel	= 1 if panel data were used (ref. cat.: time series).	0.220	0.415
Length	The logarithm of the length of the data sample used in the primary study (in years).	3.123	0.289
Midpoint	The logarithm of the mean year of the data used in the study (normalized to the earliest mean year in our sample).	2.888	0.492
<i>Specification characteristics</i>			
GDP Defl.	= 1 if GDP deflator is included in the VAR model instead of CPI.	0.069	0.254
Foreign IR	= 1 if a foreign interest rate is included.	0.026	0.158
Credit	= 1 if credit is included.	0.278	0.448
Consumption	= 1 if consumption is included.	0.277	0.448
Res. Invest	= 1 if a measure of residential investment is included.	0.175	0.380
Money Supply	= 1 if a measure of the money supply is included.	0.204	0.403
Exch. Rate	= 1 if the exchange rate is included	0.252	0.435
Long-run IR	= 1 if the long-run interest rate (in addition to the short-run interest rate) is included.	0.197	0.398
Real HP	= 1 if real instead of nominal house prices are used.	0.940	0.237
Lags	The number of lags (in quarters) included in the model.	3.320	1.531
Time Trend	= 1 if the study uses detrended data or a time trend is added to the regression.	0.367	0.482
<i>Estimation characteristics</i>			
BVAR	= 1 if a Bayesian VAR model is employed in the primary study.	0.152	0.359
Sign Restr. HP	= 1 if sign restrictions are used in the VAR model and are imposed on the house price variable (ref. cat.: Cholesky decomposition).	0.048	0.214
Sign Restr. Other	= 1 if sign restrictions are used in the VAR model but are not imposed on the house price variable (ref. cat.: Cholesky decomposition).	0.054	0.226
Nonrecursive	= 1 if another nonrecursive identification is used in the VAR model (ref. cat.: Cholesky decomposition).	0.142	0.349
<i>Publication characteristics</i>			

Continued on next page

Table C1: Description and summary statistics of regression variables (Continued)

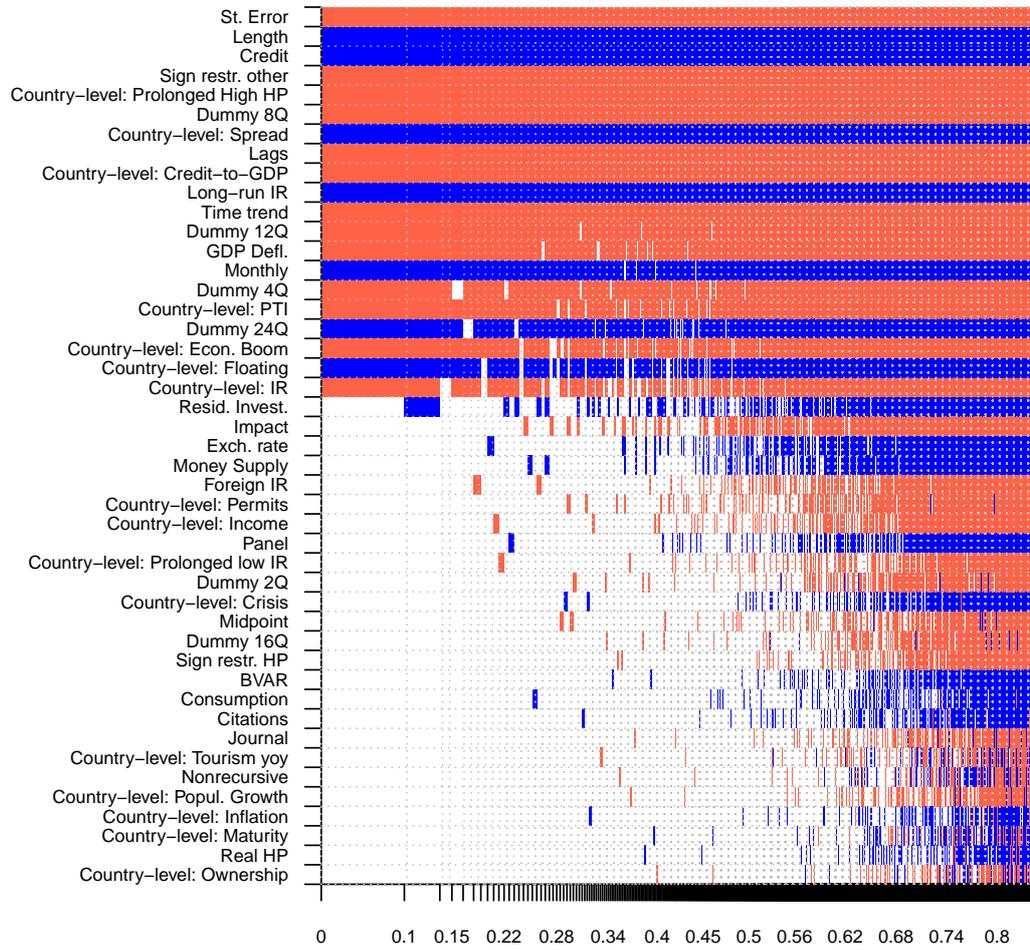
Label	Description	Mean	SD
Citations	The logarithm of the mean number of annual citations of the study received during the first three years after its publication.	2.982	1.003
Impact Journal	The recursive discounted RePEc impact factor of the outlet. = 1 if the study is published in a peer-reviewed journal.	0.519 0.412	0.556 0.492
<i>Structural heterogeneity (computed for the period and country for which the VAR is estimated)</i>			
Crisis	The number of years (out of those used in the time span of the primary study) during which a banking crisis occurred.	3.645	2.900
IR	The average three-month interest rate, OECD.	7.082	2.471
Prolonged Low IR	The number of consecutive years (out of those used in the time span of the primary study) during which the short-run interest rate was below its long-run average.	9.265	5.376
Spread	The average difference between short-term and long-term interest rates.	0.681	0.494
Floating	The share of loans with a floating interest rate.	50.698	27.273
Tourism YoY	The growth rate of the number of arrivals per capita.	3.461	4.893
Income	Average disposable income per household per capita in US dollars, OECD.	9.772	0.357
Inflation	Average consumer price inflation, OECD.	4.163	2.318
Credit-to-GDP	The credit-to-GDP ratio, BIS.	126.545	33.323
Popul. Growth	Average annual population growth, World Bank.	0.616	0.393
PTI	The standardized price-to-income ratio for the housing market.	94.527	8.776
Prolonged High HP	The number of periods (out of those used in the time span of the primary study) with above-average house price growth.	12.414	4.614
Permits	The number of building permits issued in comparison to its long-run average.	101.074	21.868
Maturity	The average maturity of mortgage loans in logs.	3.088	0.224
Ownership	The share of home ownership.	61.205	8.889
Econ. Boom	The number of periods (out of those used in the time span of the primary study) with a positive output gap.	5.641	3.367

Table C2: Results of frequentist model averaging

Category	Variable	Coef.	Std. Er.	<i>p</i> -value
<i>Publication bias</i>	SE	-1.633	0.186	0.000
<i>Data characteristics</i>	Monthly	0.361	0.470	0.442
	Panel	0.000	0.184	1.000
	Length	1.493	0.815	0.067
	Midpoint	0.110	0.327	0.736
<i>Specification characteristics</i>	GDP Defl.	-0.202	0.287	0.481
	Foreign IR	-0.244	0.399	0.541
	<i>Credit</i>	<i>0.264</i>	<i>0.205</i>	<i>0.197</i>
	Consumption	-0.128	0.247	0.606
	Resid. Invest.	0.181	0.310	0.559
	Money Supply	0.340	0.326	0.297
	Exch. rate	0.165	0.224	0.462
	Long-run IR	0.196	0.251	0.435
	Real HP	0.285	0.347	0.410
	Lags	-0.048	0.078	0.541
	Time trend	-0.205	0.212	0.334
<i>Estimation characteristics</i>	BVAR	0.398	0.400	0.319
	Sign restr. HP	-0.500	0.445	0.261
	Sign restr. other	-0.486	0.426	0.254
	Nonrecursive	0.000	0.278	1.000
<i>Publication characteristics</i>	Citations	0.026	0.112	0.817
	Impact	-0.112	0.149	0.454
	Journal	-0.024	0.198	0.903
<i>Structural heterogeneity</i>	Country-level: Crisis	0.029	0.044	0.504
	Country-level: IR	-0.126	0.103	0.221
	Country-level: Prolonged low IR	-0.029	0.032	0.363
	Country-level: Spread	0.224	0.218	0.304
	Country-level: Floating	0.005	0.004	0.285
	Country-level: Tourism yoy	-0.011	0.016	0.486
	Country-level: Income	0.257	0.568	0.651
	Country-level: Inflation	0.026	0.057	0.650
	<i>Country-level: Credit-to-GDP</i>	<i>0.013</i>	<i>0.008</i>	<i>0.103</i>
	Country-level: Popul. Growth	0.302	0.435	0.489
	<i>Country-level: PTI</i>	<i>-0.029</i>	<i>0.020</i>	<i>0.146</i>
	Country-level: Prolonged High HP	-0.070	0.042	0.097
	Country-level: Permits	0.004	0.006	0.542
	Country-level: Maturity	-0.202	0.397	0.612
	Country-level: Ownership	-0.016	0.019	0.407
	Country-level: Econ. Boom	-0.034	0.042	0.419
	Constant	-0.5212	5.2719	0.921
Observations	225			

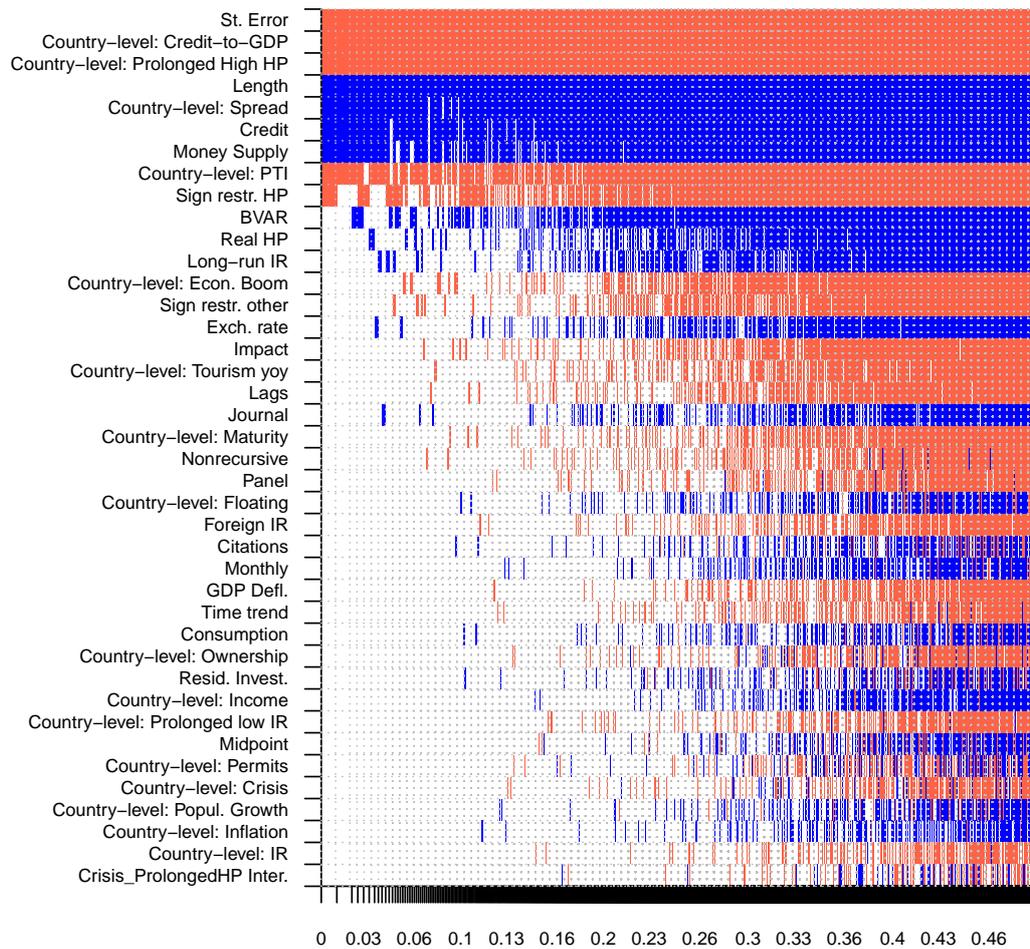
Notes: The frequentist model averaging (FMA) exercise employs Mallows' weights (Hansen, 2007) and the orthogonalization of the covariate space suggested by Amini & Parmeter (2012). Variables significant at the 10% level are shown in bold; variables that were important in BMA and have a *p*-value lower than 0.2 are indicated in italics.

Figure C1: Model inclusion in BMA with estimates for all horizons



Notes: Columns denote individual models; the variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities; the 10,000 best models are shown. To ensure convergence we employ 3 million iterations and 1 million burn-ins. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive, i.e., the transmission is weaker. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative, i.e., the transmission is stronger. No color = the variable is not included in the model. A detailed description of all the variables is available in Table C1.

Figure C2: BMA with an interaction between crisis and prolonged periods of high house prices



Notes: The interaction is the variable with the smallest posterior inclusion probability (at the bottom of the figure). Columns denote individual models; the variables are sorted by posterior inclusion probability in descending order. The horizontal axis denotes the cumulative posterior model probabilities; the 10,000 best models are shown. To ensure convergence we employ 3 million iterations and 1 million burn-ins. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive, i.e., the transmission is weaker. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative, i.e., the transmission is stronger. No color = the variable is not included in the model. A detailed description of all the variables is available in Table C1.

Table C3: A robustness check using ordinary least squares

Category	Variable	1Q	2Q	4Q	8Q	12Q	16Q
<i>Publication bias</i>	SE	-0.764* (0.395)	-1.251*** (0.271)	-1.681*** (0.244)	-1.618*** (0.133)	-0.953*** (0.230)	-0.521*** (0.148)
<i>Data characteristics</i>	Panel	-0.263** (0.109)	-0.188 (0.131)	-0.00259 (0.158)	0.0329 (0.241)	0.349 (0.293)	0.576** (0.289)
	Length	1.428** (0.706)	1.437** (0.574)	1.788** (0.730)	2.571*** (0.688)	1.990*** (0.526)	0.918*** (0.346)
<i>Specification characteristics</i>	GDP Deflator	0.0474 (0.191)	0.164 (0.187)	-0.328** (0.158)	-0.696*** (0.209)	-0.831*** (0.252)	-0.848*** (0.275)
	Credit	0.0321 (0.102)	0.223*** (0.0418)	0.370*** (0.117)	0.597*** (0.161)	0.440*** (0.134)	0.330*** (0.123)
	Resid. invest.	0.648** (0.289)	0.426* (0.257)	0.205 (0.227)	0.113 (0.315)	0.286 (0.325)	0.442 (0.322)
	Money Supply	0.0259 (0.140)	0.00278 (0.176)	0.615*** (0.230)	1.005*** (0.239)	0.660** (0.288)	0.173 (0.279)
	Exchange rate	-0.111 (0.127)	0.0978 (0.170)	0.214 (0.152)	0.254* (0.138)	0.533*** (0.162)	0.517*** (0.168)
	Long-run IR	-0.311* (0.177)	-0.0493 (0.162)	0.266 (0.214)	0.455* (0.234)	0.321* (0.168)	0.129 (0.170)
	Lags	-0.0418 (0.0378)	-0.0480 (0.0373)	-0.0200 (0.0317)	-0.0982* (0.0558)	-0.138*** (0.0457)	-0.186*** (0.0430)
	Time trend	-0.00914 (0.132)	-0.151 (0.147)	-0.253** (0.109)	-0.637*** (0.202)	-0.458*** (0.156)	-0.334** (0.148)
<i>Estimation characteristics</i>	BVAR	0.0134 (0.295)	0.134 (0.349)	0.497 (0.477)	0.732* (0.444)	0.281 (0.426)	0.0655 (0.346)
	Sign restr. other	0.597 (0.402)	0.259 (0.338)	-0.594 (0.402)	-1.359*** (0.455)	-1.685*** (0.527)	-1.667*** (0.515)
	Nonrecursive	0.285 (0.230)	0.218 (0.272)	-0.351 (0.242)	-0.682*** (0.253)	-0.0236 (0.237)	0.560*** (0.177)
<i>Publication characteristics</i>	Citations	0.0138 (0.0911)	-0.0458 (0.0710)	0.0690 (0.0766)	0.247** (0.114)	0.0308 (0.0939)	-0.176** (0.0810)
	Impact	-0.120 (0.141)	-0.0670 (0.101)	-0.147** (0.0671)	-0.311** (0.126)	-0.613*** (0.0973)	-0.535*** (0.0322)
<i>Structural heterogeneity</i>	Country-level: IR	-0.207** (0.0813)	-0.214** (0.108)	-0.177* (0.103)	-0.187*** (0.0577)	-0.149** (0.0601)	-0.186*** (0.0429)
	Country-level: Prolonged low IR	-0.00591 (0.0230)	-0.0166 (0.0332)	-0.0405 (0.0343)	-0.0694** (0.0287)	-0.0234 (0.0267)	0.00134 (0.0301)
	Country-level: Spread	-0.204 (0.139)	-0.117 (0.198)	0.291* (0.163)	0.504*** (0.110)	0.505*** (0.121)	0.416*** (0.150)
	Country-level: Floating	0.00880*** (0.00205)	0.00441 (0.00430)	0.00520 (0.00474)	0.00803*** (0.00268)	0.00729** (0.00285)	0.00870*** (0.00292)
	Country-level: Credit-to-GDP	-0.0105* (0.00621)	-0.0128* (0.00737)	-0.0177** (0.00845)	-0.0191*** (0.00549)	-0.0153*** (0.00267)	-0.0107*** (0.00314)
	Country-level: Maturity	0.743** (0.337)	0.0517 (0.691)	-0.211 (0.646)	-0.215 (0.465)	0.381 (0.316)	0.737*** (0.283)
	Country-level: PTI	-0.0298* (0.0170)	-0.0271 (0.0211)	-0.0409*** (0.0149)	-0.0490*** (0.00691)	-0.0228** (0.0104)	-0.0192** (0.00828)
	Country-level: Prolonged High HP	-0.0620 (0.0451)	-0.0766** (0.0333)	-0.0833** (0.0400)	-0.103*** (0.0393)	-0.108*** (0.0205)	-0.0849*** (0.0251)
	Country-level: Econ. Boom	0.014 (0.027)	-0.020 (0.027)	-0.057* (0.033)	-0.074** (0.034)	-0.055* (0.033)	-0.053 (0.043)
	Observations	210	215	225	225	220	218

Notes: Standard errors, clustered at the study level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. For ease of exposition, variables which are not significant at any horizon are excluded from the table.