Empir Econ
DOI 10.1007/s00181-013-0694-5

Still puzzling: evaluating the price puzzle in an empirically identified structural vector autoregression

Selva Demiralp · Kevin D. Hoover · Stephen J. Perez

Received: 19 October 2011 / Accepted: 27 November 2012

Abstract The price puzzle, an increase in the price level associated with a contractionary monetary shock, is investigated in a rich, 12-variable SVAR in which various factors that have been mooted as solutions are considered jointly. SVARs for the pre-1980 and post-1990 periods are identified empirically using a graph-theoretic causal search algorithm combined with formal tests of the implied overidentifying restrictions. In this SVAR, the pre-1980 price puzzle depends on the characterization of monetary policy, and the post-1990 price puzzle is statistically insignificant. Commonly suggested theoretical resolutions to the price puzzle are shown to have causal implications inconsistent with the data.

Keywords Price puzzle · Monetary policy · Graph theory · Causal search · Output gap · Transmission mechanism

S. Demiralp
Department of Economics, Koç University, Sariyer, Istanbul 34450, Turkey
e-mail: sdemiralp@ku.edu.tr

K. D. Hoover (✉)
Department of Economics, Duke University, Box 90097, Durham, NC 27708-0097, USA
e-mail: kd.hoover@duke.edu
URL: www.econ.duke.edu/∼kdh9

K. D. Hoover
Department of Philosophy, Duke University, Box 900975, Durham, NC 27708-0097, USA

S. J. Perez
Department of Economics, California State University, Sacramento,
6000 J. Street, Sacramento, CA 95819-6082, USA
e-mail: sjperez@csus.edu
URL: http://www.csus.edu/indiv/p/perezs/

Published online: 02 March 2013

JEL Classification C32 · C51 · E32 · E44 · E51

1 The price puzzle and structural identification

The term “price puzzle” was coined by Eichenbaum (1992) in a commentary on Christopher Sims’s (1992) study of the effects of monetary policy in several countries. The price puzzle is the association in a structural vector autoregression (SVAR) of a contractionary shock to monetary policy (typically, but not necessarily, indicated by a positive shock to the Federal funds rate) with persistent increases in the price level. Sims had offered an explanation of the “perverse” price response:

> Policy authorities might know that inflationary pressure is about to arrive and contract to dampen the effects of these pressures. Then prices would rise after the monetary contraction (though by less than they would have without the contraction) and output would fall because of the standard effects of nominal demand contraction on real output. (Sims 1992, pp. 988–989)

To capture the response of the policy maker to anticipated inflation, Sims proposed adding an index of sensitive commodity prices into the SVAR system. While including commodity prices helped to reduce the counterintuitive price response, he noted that it does not eliminate it, which is “an embarrassment to monetarist/ISLM interpretation” (Sims 1992, p. 995). The two impulse-response functions of the consumer price index (*CPI*) to the Federal funds rate (*FFR*) in Fig. 1 illustrate the price puzzle, as well as the partial success of commodity prices in resolving it. The solid black curve displays the impulse responses from a three-variable SVAR with a Choleski order of industrial production (*INDPRO*), *CPI*, and *FFR*; while the solid gray curve displays the impulse

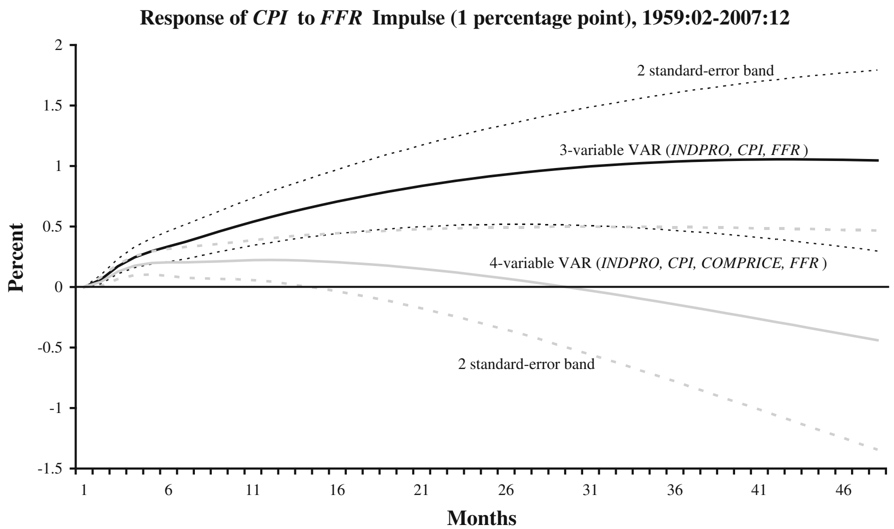


Fig. 1 Response of *CPI* to *FFR* Impulse (1 % point), 1959:02-2007:12

responses from a four-variable SVAR in which an index of sensitive commodity prices (*COMPRICE*) has been ordered ahead of *FFR*.[1]

While it has become a common practice to include commodity prices in SVAR models of the macro/monetary system, its theoretical justification is relatively weak. And, at best, it mitigates the price puzzle rather than eliminating it or generating the negative price response that most monetary economists would expect. These results proved to be robust across different studies, using different variable sets, different identification assumptions, and different sample periods (see, for example, Sims 1992; Leeper et al. 1996; Christiano et al. 1996, 1999). Some researchers interpret these results as pointing to a negligible influence of monetary policy shocks on the variation in prices (e.g., Leeper et al. 1996; Christiano et al. 2005). Other research focused on the role of monetary policy itself and whether it is adequately characterized by the Federal funds rate (Eichenbaum 1992; Strongin 1995; Bernanke and Mihov 1998; Leeper and Roush 2003). A third line of research interpreted the price puzzle as an outcome of the cost channel rather than as the result of specification error (Barth and Ramey 2001).

The large literature that has developed around the price puzzle has explored a variety of data sets and theoretical bases for identifying the appropriate SVAR. These studies have in common that they rely on a priori assumptions to identify the appropriate SVAR. In most cases, researchers restrict themselves to relatively casual discussions of appropriate recursive (Choleski) orderings for the SVAR, although some appeal to more formal theoretical considerations, and some specify nonrecursive systems. Just as Sims (1980) was concerned about the "incredible" identifying assumptions of structural macroeconometric models, we question the credibility of these approaches to identifying the SVAR. And we offer a constructive alternative.

Conditional independence relations among the data themselves typically reduce the class of admissible causal orders—in some cases to a unique order. While just-identifying restrictions impose no restrictions on the reduced form or likelihood function, overidentifying restrictions do.[2] The importance of this fact for inferring causal order from empirical data was first recognized outside of economics and has developed into a large literature known as the *graph-theoretic approach* to causal inference (e.g., see Spirtes et al. 2001; Pearl 2000). Swanson and Granger (1997) first applied these methods to the problem of determining the order of the SVAR. Subsequently, a number of economists and econometricians have further developed the methods and applied them to a variety of economic problems.[3]

[1] The order is suggested by Fig. 1 of Brissimis and Magginas (2006, p. 1230), but is for a longer time period: 1959:08 to 2007:12. For definitions of the data, see Appendix.

[2] In his classic treatment of the identification problem, Fisher (1966, p. 46) writes:

> [in the just-identified case] the equations to be solved have an infinite number of solutions; many other equations than the true structural one satisfy them. This is true no matter how large a sample we consider and no matter how we abstract from problems of sampling error. In the overidentified case, on the other hand, no problem would arise if we could so abstract… It is thus … a practical problem … rather than a problem of principle.

[3] Swanson and Granger (1997), Akleman et al. (1999) , Bessler and Loper (2001), Bessler and Lee (2002), Demiralp and Hoover (2003), Haigh et al. (2004), Awokuse (2005), Hoover et al. (2009),

Our approach will be to apply graph-theoretic methods to identify empirically an SVAR in a rich data set in which many of the solutions to the price puzzle offered in the literature can be studied. We address three key questions: First, whether the price puzzle is still a puzzle in an SVAR that respects previously unacknowledged constraints in the data? Second, whether the results are robust over time? And third, to what degree various "fixes" for the price puzzle that have been offered are compatible with such an empirically identified SVAR.

It is critical to understand that we are not offering a theoretical explanation of the price puzzle and our success or failure should not be judged on that basis. Rather we are offering, first, a characterization of key features of the data for which any successful theoretical model ought to supply an account, and, second, we are conducting a test of the ability of existing theoretical accounts to match those key features. In spirit, our investigation is no different than one in which theoretical models are tested for their ability to capture serial-correlation properties or covariances known from empirical studies to characterize the data. Our innovation is to focus on constraints implied by the data for contemporaneous causal order of the SVAR. Many economists have previously supposed that the data do not constrain the contemporaneous causal order and that, therefore, the only preferences one might have must derive from a priori theory. But as has been more and more widely acknowledged, research on graph-theoretic causal search demonstrates that this supposition is false: data imply—sometimes univocally, sometimes not—that some theoretical presuppositions about contemporaneous causal structure are compatible with the data and some are not. We thus have the basis for discriminating among those models according to the assumptions that they make about causal order.

2 Empirical identification of the SVAR

There are two relevant senses of *identification* in play in the specification of a SVAR. To understand the first sense, start with an SVAR written as:

$$\mathbf{A}_0\mathbf{Y}_t = \mathbf{A}(L)\mathbf{Y}_{t-1} + \mathbf{E}_t, \tag{1}$$

where $\mathbf{Y}_t$ is an $n \times 1$ vector of contemporaneous variables, $\mathbf{A}_0$ is a full rank $n \times n$ matrix with ones on the main diagonal and possibly non-zero off-diagonal elements; $\mathbf{A}(L)$ is a polynomial in the lag operator, L; and $\mathbf{E}_t$ is an $n \times 1$ vector of identical, independent, normal error terms $\mathbf{E}_t \sim N(0, \mathbf{\Sigma})$. Let $\mathbf{E} = [\mathbf{E}_t]$, $t = 1, 2, \ldots, T$, then the covariance matrix $\mathbf{\Sigma} = E(\mathbf{E}\mathbf{E}')$ is diagonal. The individual error terms (shocks) can be assigned unequivocally to particular equations because $\mathbf{\Sigma}$ is diagonal. The matrix $\mathbf{A}_0$ defines the causal interrelationships among the contemporaneous variables. The system is *identified* in this sense of delivering *unique* coefficient estimates only if there are $n(n-1)/2$ zero restrictions on $\mathbf{A}_0$ (see Hendry et al. 2009 for a discussion of the meaning of identification). For any just-identified system, $\mathbf{A}_0$ can be rendered

Footnote 3 continued
and Reale and Tunnicliffe Wilson (2001, 2002) offer a substantially different, but still graph-theoretically based, method approach to the causal ordering of SVARs. A comparison of these methods with the PC algorithm would be interesting but beyond the scope of this article.

lower triangular by selecting the appropriate order of the variables in $\mathbf{Y}$ along with the conformable order for the rows of $\mathbf{A}_0$. This is the *recursive* (or *Wold causal)order* (Wold 1954).

There is, however, an economically more central sense of identification. Starting with the SVAR that actually corresponds to the data-generating process (DGP), premultiplying by $\mathbf{A}_0^{-1}$ yields the reduced-form or ordinary vector autoregression (VAR):

$$\mathbf{Y}_t = \mathbf{A}_0^{-1}\mathbf{A}(L)\mathbf{Y}_{t-1} + \mathbf{A}_0^{-1}\mathbf{E}_t = \mathbf{B}(L)\mathbf{Y}_{t-1} + \mathbf{U}_t. \tag{2}$$

The transformed error terms are $\mathbf{U} = [\mathbf{U}_t],\ t = 1, 2, \ldots, T$. And, in general, $\boldsymbol{\Omega} = E(\mathbf{U}\mathbf{U}')$ is, unlike $\boldsymbol{\Sigma}$, not diagonal. If we *knew* $\mathbf{A}_0$, then recovery of the SVAR (1) from the easily estimated VAR (2) would be straightforward. There are, however, a large number of $n \times n$ matrices, $\mathbf{P}_i^{-1}$ that may be used to premultiply Eq. (2) such that the resulting covariance matrix, $E(\mathbf{P}_i^{-1}\mathbf{U}(\mathbf{P}_i^{-1}\mathbf{U})')$, is diagonal. The central problem of identification in the sense of *correspondence to the desired entity* is to choose the one member of $\mathbf{P}_i$ that recapitulates the data-generating process—that is to find $\mathbf{P}_i = \mathbf{A}_0$, when $\mathbf{A}_0$ is unknown (see Hendry et al. 2009).

If we restrict ourselves to zero restrictions on recursive systems, then any just-identified $\mathbf{P}_i$ can be arranged in the form of one of the $n!$ Choleski orderings (or decompositions) corresponding to the each of the possible permutations of the variables in $\mathbf{Y}$. For a given permutation, the Choleski ordering is the unique lower triangular $\mathbf{P}_i$ such that $\mathbf{P}_i\boldsymbol{\Omega}\mathbf{P}_i'$ is diagonal. The just-identified SVARs specified by each of the Choleski orderings are observationally equivalent in the sense that each has the same likelihood. Yet, the different orderings have different consequences for the dynamics of the SVAR. If we are lucky or if we have additional knowledge of the right kind, so that we know that the DGP has a Choleski order and we know which one, then $\mathbf{P}_i\boldsymbol{\Omega}\mathbf{P}_i' = \boldsymbol{\Sigma},\ \mathbf{P}_i = \mathbf{A}_0$ and we can recover the true SVAR.

Since we can neither count on luck nor count on recognizing it when we have it, on the basis of what knowledge should we choose a particular $\mathbf{P}_i$? If we are restricted to just-identified SVARs, then we have little choice but to appeal to economic theory to tell us what the causal order should be. This is, in fact, what investigators of the price puzzle have typically claimed to do—and indeed what almost all practitioners of VAR methodologies profess to do. But since formal economic theory is rarely decisive about causal order, VAR practitioners either choose the order based on some common-sense criteria or even arbitrarily.

If, however, the true SVAR is overidentified, a graph-theoretic approach offers another option. Spirtes et al. (2001) and Pearl (2000) represent causal relationships with a causal graph in which arrows connect causal variables to their effects. They show that there are mappings between graphs and the probability distributions of variables. In particular, certain graphical patterns (causal relationships) imply particular relationships of conditional independence and dependence among the variables. The graph of the data-generating process can also be represented through the restrictions on $\mathbf{A}_0$. Each arrowhead corresponds to a nonzero coefficient in $\mathbf{A}_0$. Working backwards from statistical measures of conditional independence and dependence, it is possible to infer an equivalence class of graphs compatible with the data. Sometimes that equivalence class has only a single member, and then $\mathbf{A}_0$ can be identified statistically. But even when the resulting graph is not fully identified, it will normally narrow the equivalence class to a workable number of alternatives far smaller than the number of possible identified models.[4]

Restricting ourselves to recursive orderings (or *directed acyclical graphs*—sometimes referred to as *DAG*s), the key ideas of the graph-theoretic approach are simple.[5] Suppose that $A \rightarrow B \rightarrow C$ (that is, A causes B causes C). A and C would be dependent, but conditional on B, they would be independent. Similarly for $A \leftarrow B \leftarrow C$. In each case, B is said to *screen off* A from C. Suppose that $A \leftarrow B \rightarrow C$. Then, once again A and C would be dependent, but conditional on B, they would be independent. B is said to be the *common cause* of A and C. Now suppose that A and B are independent conditional on sets of variables that exclude C or its descendants, and $A \rightarrow C \leftarrow B$, and none of the variables that cause A or B directly causes C. Then, conditional on C, A, and B are dependent. C is called an *unshielded collider* on the path *ACB*. (A *shielded* collider would have a direct link between A and B.)

Causal search algorithms use a statistical measure of independence, commonly a measure of conditional correlation, to check systematically the patterns of conditional independence and dependence and to work backwards to the class of admissible causal structures. In this paper, we use the PC algorithm, the most common of the causal-search algorithms (Spirtes et al. 2001, pp. 84–85, Pearl 2000, pp. 49–51, Cooper 1999, p. 45). It assumes that graphs are *acylical* or strictly recursive—that is, loops in which $A \rightarrow B \rightarrow C \rightarrow A$ are ruled out. Naturally, acyclicality also rules out simultaneity—that is, a very tight loop in which $A \rightarrow B \rightarrow A$ (or $A \leftrightarrow B$). While the assumption of acyclicality is restrictive, it is nonetheless less restrictive than limiting SVARs to Choleski orders. The Choleski orders, typically taken as the default in VAR studies, assume that all the lower triangular elements of the $\mathbf{A}_0$ matrix are nonzero, while our assumption will be that they may or may not be nonzero, so that Choleski orders are a particular subset of acyclical orders.

We provide only an informal discussion of the PC algorithm here (see Cooper 1999; Demiralp and Hoover 2003 for more detailed accounts). Essentially, it begins with the complete set of variables in the VAR densely connected by undirected edges, represented as lines in a graph without arrowheads.[6] It then tests for unconditional

[4] The number of acyclical models is huge. An estimate can be formed by considering that each is nested in a just-identified model. If there are n variables, then there are $n!$ such models. Consider the $\mathbf{A}_0$ matrix for each such model with the variables ordered such that there are zeroes above the main diagonal. Then every possible acyclical model nested in that just-identified model places either places a zero or a nonzero value in the cells below the main diagonal. There are $n(n-1)/2$ such cells, so there are $2^{n(n-1)/2}$ combinations for each just-identified model. For the 12-variable systems reported below, an upper bound is given by $12!(2^{66}) = 3.53 \times 10^{28}$ models. This is an overestimate of the number of models, since some overidentified models are nested in more than one just-identified model and are counted more than once. But a lower bound is the number of combinations nested in a single just-identified model ($2^{66} = 7.38 \times 10^{19}$)—still a huge number.

[5] More formal and more detailed expositions of these ideas are found in Spirtes et al. (2001) and Pearl (2000). Cooper (1999), Demiralp and Hoover (2003), and Hoover (2005) provide more concise expositions.

[6] The algorithm operates under the assumption of *causal sufficiency*—namely that any omitted variables cause at most one of the included variables and so may have their effects absorbed in the error terms and constants without altering the dependencies among other variables in the system. While this is a strong

correlations and removes any uncorrelated edges. It then tests for correlations conditional on one other variable, again removing edges for which correlations vanish on conditioning. It then repeats the procedure, conditioning on two, then three, then … up to the maximum possible number of variables. The result is an undirected *skeleton*. The algorithm begins orienting edges by seeking triples of linked variables (A — B — C) in which the variables on the endpoints (A and C) are independent on some conditioning set, but become dependent when conditioning on the intermediate variable (B). This is the pattern of an unshielded collider, and the edges are then oriented ($A \rightarrow B \leftarrow C$). Some edges may be oriented *logically* (rather than statistically), based, first, on maintaining the assumption of acyclicality and, second, on avoiding inferring unshielded colliders that were not identified on the basis of tests of conditional independence.

Not all causal graphs are recoverable from the probability distribution. Graphs that have the same unshielded colliders and the same skeleton are observationally equivalent (Pearl 2000, Theorem 1.2.8, p. 19). If the true graph is a member of an observationally equivalent set, the algorithm will not orient the edges that distinguish one member of the set from another. In these cases, unoriented edges may be oriented in either direction without changing the likelihood, provided that no new unshielded colliders or cyclicality is introduced. Also, the maintained assumption of acyclicality notwithstanding, the algorithm will sometimes identify edges as bidirectional as a result of either ambiguity in the statistical test because of small samples.

Following Swanson and Granger (1997), we treat the estimated errors ($\hat{\mathbf{U}}_t$) from the VAR in Eq. (2) as the original data purged of their dynamics (an application of the Frisch–Waugh theorem; see Juselius 2006, p. 116). The covariance matrix of these transformed data ($\hat{\Omega}$) provides the necessary information for computing the various conditional correlations required by the PC algorithm. The algorithm selects a graph that best represents the causal order, and this graph in turn corresponds to particular zeroes in (and overidentifying restrictions on) $\mathbf{A}_0$.[7]

The procedure constructs many simulations of the VAR, Eq. (2), based on the actual coefficient estimates ($\hat{\mathbf{B}}(L)$) and resampling $\hat{\mathbf{U}}_t$ by columns (to preserve the contemporaneous structure of interdependence among the variables), runs the search algorithm, and keeps track of the distribution of edges in the resulting graphs. The bootstrap method is essentially heuristic and provides guidance for more formal investigation based on testing the overidentifying restrictions on $\mathbf{A}_0$.

Footnote 6 continued
assumption, it is no stronger than what is normally made with any SVAR model. Spirtes et al. (2001) have developed algorithms that relax this assumption and detect the effect of omitted latent variables on the independence relations among the included variables.

[7] Demiralp and Hoover (2003) provide Monte Carlo evidence that shows that the PC algorithm is highly effective at recovering the skeleton of the graph of the data-generating process and moderately effective at recovering the directions of individual links, provided that signal-to-noise ratios are high enough. Demiralp et al. (2008) develop and validate a bootstrap procedure to assess the effectiveness of the closely related SGS algorithm.

3 The data

To analyze the price puzzle we use a relatively rich data set. The data consist of 12 monthly series for the United States that run from 1959:02 to 2007:06. Sources and details are provided on Hoover's website (http://public.econ.duke.edu/~kdh9/).

Monetary policy is represented both by the Federal funds rate (*FFR*) and two reserve components: (the logarithms of) borrowed reserves (*BORRES*) and nonborrowed reserves (*NBORRES*). Financial markets are represented by two monetary aggregates (the logarithms of) $M1$ and (the non-M1 components of) M, as well as by three interest rates: the own-rate of interest on M2 (*M2OWN*), the 3-month Treasury bill rate ($R3M$), and the 10-year Treasury bond rate ($R10Y$). Prices are represented by (the logarithms of) the consumer price index (*CPI*) and an index of sensitive commodity prices (*COMPRICE*). Finally, the real economy is represented by the (logarithm of) industrial production (*INDPRO*) and the output gap (*GAP*).

4 Identifying the structural vector autoregression

Since U.S. monetary policy—and perhaps the real economy itself—falls into distinct regimes, it is unlikely that the same SVAR characterizes the economy over the past 50 years. Therefore, we first identify SVARs for two periods.

The first starts at the beginning of our data set in 1959:02 and runs through 1979:09, the last month before the famous new reserve-targeting operating procedures adopted by the Federal Reserve under chairman Paul Volcker on 6 October 1979.

While reserve targeting was abandoned between 1982 and 1984, it took some time for monetary policy to stabilize. The second period, therefore, begins in 1990:01, and ends in 2007:06. There is no general agreement about exactly when the Federal Reserve abandoned the post-1984 policy of targeting borrowed reserves; it made no formal announcement. Meulendyke (1998) dates the transition to the stock-market crash of 1987, while Hamilton and Jorda (2002) place it at November 1989. It is generally agreed that the new policy was in place by the beginning of 1990. The end date is set to end ahead of the development of the monetary conditions that led to the introduction of the Term Auction Facility in August 2007. The new facility essentially allows banks to borrow from the Federal Reserve any quantity of reserves they desire provided that they have enough collateral. One consequence is that nonborrowed reserves have been allowed to become negative.[8]

4.1 The earlier period: before October 1979

To identify the contemporaneous causal order among our variables, we first estimate the unrestricted VAR—that is, Eq. (2), where $\mathbf{Y}_t = [FFR, BORRES, NBORRES, M1, M2, M2OWN, R3M, R10Y, CPI, COMPRICE, INDPRO, GAP]'$.[9] The corrected Akaike

[8] See http://www.federalreserve.gov/monetarypolicy/taf.htm.

[9] Estimates were conducted using *E-views 5.1* econometric software. The actual VAR and final SVAR estimates, as well as tests for VAR lag length, are available on the Hoover's website: http://public.econ.duke.edu/~kdh9/.

Information Criterion of Hurvich and Tsai (1991), which seems well-adapted to this problem, selects a lag length of one. Both the Hannan–Quinn and Schwarz criteria also select only a single lag. We estimate the VAR, however, with a lag length of two in order to allow more flexibility—especially to capture both inflation and acceleration-of-inflation terms in price equations. The covariance matrix of $\hat{\mathbf{U}}_t$ serves as the input into the PC algorithm. The critical value for the tests of conditional correlation used in the algorithm is set at 10 %, as suggested on the basis of Monte Carlo studies by Scheines et al. (1994, pp. 103–107) for the number of available observations (246 after accounting for lags). A critical value higher than the more common 5 % is also justified by our concern not to restrict the specification too much. Here a higher rate of type-II error biases us in favor of finding causal connections at the risk that they are not in the data-generating process. We will subsequently investigate their inclusion using tests of overidentifying restrictions.

The first three columns of Table 1 show the graphical structure of the VAR identified by the PC algorithm applied to the VAR.[10] To evaluate the reliability of this identification, we apply the bootstrap procedure of Demiralp et al. (2008) with 10,000 replications. There are 66 possible edges among twelve variables. Table 1 shows the results of the PC algorithm and the bootstrap procedure for 19 of them—the 13 selected by the algorithm plus all the edges selected in more than 10 % of the bootstrap replications. The next five columns show the actual distribution of edges from the bootstrap. The last three columns present summary statistics: *exists* is the fraction of replications in which an edge is found (100 − no edge); *directed* is the percentage of edges discovered that have a definite direction; *net direction* is the difference between the percentage of edges directed rightward (→) and edges directed leftward (←).

Three of the identified edges are undirected (—); while, despite the assumption of acyclicality built into the algorithm, five are shown as bidirectional (↔), which as we noted already could be the result of low power when testing for independence is conducted in a piecewise manner.[11]

The orientation of undirected edges does not affect the likelihood. It also does not much affect the impulse-response function of *CPI* to *FFR*, which is the focus of our investigation. We set these edges to their dominant one-way orientation according to the *net direction* column of Table 1. Our initial inclination is to accept provisionally the bidirectional edges as bidirectional and then to use tests of overidentification to

[10] Implementation of the PC algorithm and the bootstrap evaluations use a *Gauss* program downloadable from the Hoover's website: http://public.econ.duke.edu/~kdh9/.

[11] By "piecewise" we mean that the algorithm tests pairs and triples of variables in isolation but does not ever test all the variables jointly. To illustrate how low power could result in a bidirectional edge, consider the following case in which our focus is on the bidirectional edge between *FFR* and *CPI*: A subgraph of the graph represented in Table 1 is $NBORRES \rightarrow FFR \leftrightarrow CPI \leftrightarrow COMP$. Starting from the skeleton in which all of the arrows lack heads, in order to orient the first three variables as $NBORRES \rightarrow FFR \leftarrow CPI$, we must not reject the null $\rho_{NBORRES,CPI} = 0$, where ρ is the correlation coefficient, but we must reject the null $\rho_{NBORRES,\,CPI|FFR} = 0$, where the vertical (|) indicates conditioning. Suppose that we have done that successfully. To orient $FFR \rightarrow CPI \leftarrow COMP$, which would turn the previous single-headed arrow between *FFR* and *CPI* into a two-headed arrow, we must not reject the null $\rho_{FFR,\,COMP} = 0$, where ρ is the correlation coefficient, but we must reject the null $\rho_{FFR,\,COMP|CPI} = 0$. If our first orientation was correct, then this should not be possible, since *COMP* causes *CPI* and *CPI* causes *FFR*. It could, however, happen if we made a type II error: our null $\rho_{FFR,\,COMP} = 0$ was in fact false and yet we accepted it.

Table 1 Bootstrap evaluation of the initial causal graph, 1959:02-1979:09

Causal order selected by the PC algorithm[a]			Edge identification (percent of bootstrap realizations)[b]					Summary statistics for bootstrap distribution[c]		
			—	←	No edge	→	↔	Exists	Directed	Net direction
INDPRO	—	*GAP*	85	6	0	8	1	100	15	2
R3M	←	*R10Y*	7	22	0	39	32	100	93	17
BORRES	—	*NBORRES*	41	34	0	18	7	100	59	−16
BORRES	→	*FFR*	22	24	2	40	13	98	78	16
FFR	↔	*R3M*	13	33	10	16	28	90	86	−16
M1	—	*M2*	28	17	32	15	8	68	59	−2
R10Y	←	*M2OWN*	3	27	42	4	24	58	95	−23
M2OWN	↔	*COMPRICE*	8	18	55	5	14	45	82	−13
R3M	↔	*M2OWN*	0	17	64	1	18	36	99	−16
FFR	↔	*CPI*	0	20	66	0	14	34	100	−20
NBORRES	→	*FFR*	2	1	67	25	5	33	93	24
BORRES	→	*R3M*	1	1	73	12	13	27	97	11
CPI	↔	*COMPRICE*	4	5	74	10	8	26	84	5
R10Y	No edge	*COMPRICE*	0	11	82	0	7	18	100	−11
BORRES	No edge	*M1*	0	5	84	0	11	16	100	−5
R10Y	No edge	*M2*	0	4	85	0	11	15	100	−4
M2OWN	No edge	*M1*	1	4	86	1	9	14	96	−3
R10Y	No edge	*CPI*	0	7	87	0	5	13	98	−7
NBORRES	No edge	*M1*	0	3	89	0	7	11	100	−3

[a] 19 of 66 candidate edges; only edges that are identified as existing in 10 % or more of the bootstrap replications are shown
[b] Values indicate percentage of 10,000 bootstrap replications in which each type of edge is found. Based on the procedure in Demiralp et al. (2008) with critical value of 2.5 % for tests of conditional correlation (corresponding to the 10 % critical value used in the PC algorithm)
[c] *Exists* the percentage of bootstrap replications in which an edge is selected (= 100 − no edge); *directed* edges directed as a percentage of edges selected; *net direction* difference between edges directed right (→) and left (←)

check whether each can be oriented in a single direction. However, the $\mathbf{A}_0$ matrix based on the graph in Table 1 is not identified. So, we use a modified strategy. A complete description of the investigation is reported in "Appendix".

First, we estimate an SVAR with the undirected and bidirected edges set to the dominant one-way orientation in the bootstrap replications as indicated by the *net direction* column of Table 1. The overidentifying restrictions implicit in this graph cannot be rejected against any of the just-identified (Choleski) orders in which it is nested: $\chi^2(53) = 63.48 (p = 0.15)$. Next we evaluate each bidirectional edge one at a time. Holding other edges constant, we reëstimate the SVAR with the candidate edge oriented in each of the one-way directions as well as bidirectionally, and select the orientation with the largest p value for the test of overidentifying restrictions.

Once all of the edges have been oriented in this way, we conduct an informal general-to-specific search of the resulting SVAR in the spirit of the London School of Economics (LSE) methodology of David Hendry and his colleagues, which have been incorporated into automatic model-selection programs (particularly, in the *PCGets* software and *Autometrics* software—see Hendry and Krolzig 2001; Doornik and Hendry 2007). These methods have been demonstrated to commit errors of omission at rates close to those of the size of the underlying variable exclusion tests (here 5 %) and to commit error of commission at rates similar to those of one-shot estimates when the correct specification is known in advance and no search is conducted (see Hoover and Perez 1999, 2004; Krolzig and Hendry 2001).

The search is over the structure of $\mathbf{A}_0$ matrix corresponding to the graph after the undirected and bidirected edges have been resolved as previously described. The search removes the edge (sets the relevant coefficient in the $\mathbf{A}_0$ matrix to zero) with the lowest insignificant t statistic. Each time an edge is removed, the p value of the likelihood ratio test against the just-identified model is calculated. If the implied overidentifying restrictions cannot be rejected, the coefficient of the remaining edges with the lowest t statistic is treated in the same way. If the overidentifying test rejects the exclusion or if the removal of the edge results in a failure of identification, then the edge is replaced and the search proceeds to the edge with the next lowest t statistic. The search continues until either all edges are statistically significant or no further edges can be removed without rejecting the overidentifying assumptions.

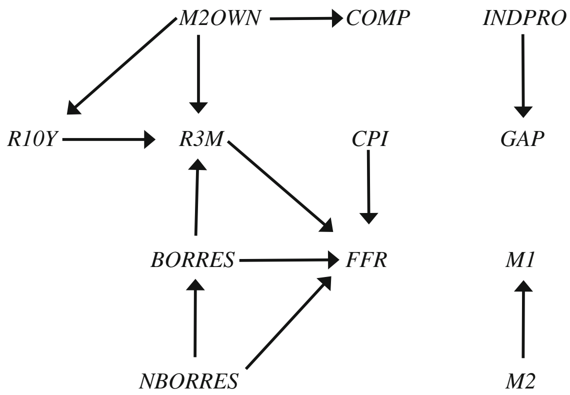
The results of the search procedure are the graph and the $\mathbf{A}_0$ matrix shown in Fig. 2. The overidentifying restrictions implied in the graph cannot be rejected at a 10 % size: $\chi^2(54) = 66.69(p = 0.12)$. The most striking thing about $\mathbf{A}_0$ matrix is that it is very parsimoniously parameterized in comparison to any Choleski ordering. The order of the variables in the $\mathbf{A}_0$ matrix was selected to render it lower triangular, so as to emphasize that the graph is overidentified and nested in the corresponding Choleski ordering. It is also nested in other Choleski orderings, but the selected graph is not nested in the vast majority of the 12! = 479,001,600 possible Choleski orderings. This illustrates one feature of the graph-theoretic approach: all Choleski orders return the same value of the likelihood function, but not all are compatible with the information about conditional independence encoded into the likelihood function.

4.2 The later period: January 1990 to June 2007

The procedures for the later period are identical to those for the earlier period. Once again, the corrected Akaike Information Criterion of Hurvich and Tsai (1991), as well as the Hannan–Quinn and Schwarz criteria, selects a lag length of one. And for the same reasons as before, we estimate a VAR with two lags. Table 2 reports the eleven edges identified by the PC algorithm plus the distribution and summary statistics for two additional edges that are selected in more bootstrap replications than the least-often selected of the original eleven edges ($CPI \rightarrow R3M$).

Contemporaneous Causal Order of the SVAR 1959:02-1979:09

The Causal Graph



The $\mathbf{A}_0$ Matrix

$$
\begin{array}{l}
NBORRES \\ M2O \\ M2 \\ CPI \\ INDPRO \\ BORRES \\ R10Y \\ M1 \\ GAP \\ COMP \\ R3M \\ FFR
\end{array}
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\alpha_{61} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \alpha_{72} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \alpha_{83} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \alpha_{95} & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & \alpha_{102} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & \alpha_{112} & 0 & 0 & 0 & \alpha_{116} & \alpha_{117} & 0 & 0 & 0 & 1 & 0 \\
\alpha_{121} & 0 & 0 & \alpha_{124} & 0 & \alpha_{126} & 0 & 0 & 0 & 0 & \alpha_{1211} & 1
\end{bmatrix}
$$

Fig. 2 Contemporaneous causal order of the SVAR 1959:02-1979:09

In this case, there is one undirected edge and four bidirectional edges. When these five edges are oriented according to the dominant direction in the last column of Table 2, the implied overidentifying restrictions cannot be rejected at 5 % critical value: $\chi^2(55) = 69(p = 0.09)$. To test the robustness of the one-way orientation of the bidirectional edges, we reverse each edge sequentially, starting with the edge with the lowest absolute value for the *net direction* statistics in Table 2, testing the implied overidentifying restrictions at each step, looking for the least restrictive set of orientations. The test of overidentifying restrictions for the final specification yields a likelihood ratio statistic $\chi^2(55) = 63(p = 0.22)$. As this specification is not rejected against the just-identified model and as all of its coefficients are statistically

Table 2 Bootstrap evaluation of the initial causal graph, 1990:01-2007:12

Causal order selected by the PC algorithm[a]			Edge identification (percent of bootstrap realizations)[b]					Summary statistics for bootstrap distribution[c]		
			—	←	No edge	→	↔	Exists	Directed	Net direction
INDPRO	←	*GAP*	58	26	0	13	4	100	42	−13
FFR	→	*R3M*	15	5	0	75	5	100	85	70
FFR	→	*M2OWN*	50	4	0	43	3	100	50	38
R3M	↔	*R10Y*	7	40	0	10	42	100	93	−29
NBORRES	→	*M1*	29	24	1	27	19	99	70	3
BORRES	—	*NBORRES*	33	13	31	19	4	69	52	6
M2OWN	↔	*M1*	0	3	60	3	34	40	99	0
R10Y	↔	*COMPRICE*	1	14	63	2	20	37	96	−12
NBORRES	→	*COMPRICE*	4	11	69	4	11	31	86	−6
R10Y	↔	*INDPRO*	1	3	73	5	18	27	96	1
BORRES	No edge	*COMPRICE*	6	2	80	9	3	20	70	6
R3M	No edge	*GAP*	1	7	82	4	7	18	95	−3
R3M	←	*CPI*	0	11	84	0	5	16	100	−11

[a] 13 of 66 candidate edges; the only edges shown are ones identified by the PC algorithm or ones for which the *exists* statistic is higher than the lowest value for an edge selected by the PC algorithm
[b] Values indicate percentage of 10,000 bootstrap replications in which each type of edge is found. Based on the procedure in Demiralp et al. (2008) with critical value of 2.5 % for tests of conditional correlation (corresponding to the 10 % critical value used in the PC algorithm)
[c] *Exists* the percentage of bootstrap replications in which an edge is selected (= 100 − no edge); *directed* edges directed as a percentage of edges selected; *net direction* difference between edges directed right (→) and left (←)

significant —the least significant among them having a p value of 0.03—there is no need to conduct a general-to-specific specification search as we did for the earlier period. The resulting graph and the $\mathbf{A}_0$ matrix are shown in Fig. 3.

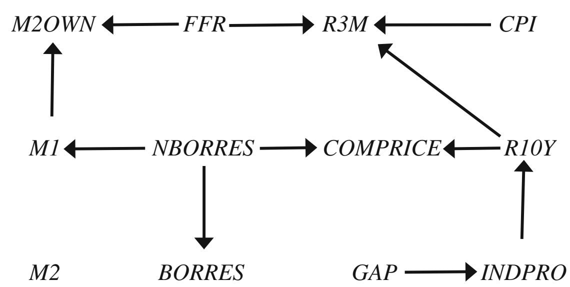
5 The price puzzle in light of the empirical identification

5.1 The persistence of the price puzzle

We now consider the light that empirical identification of the SVARs for the two periods sheds on the price puzzle. The causal structure of the two periods is very different. An indication of the change in causal structure is found in a test of the overidentifying restrictions of the earlier period on the VAR of the later period. The overidentifying restrictions are rejected: $\chi^2(54) = 244(p = 0.00)$. Similarly, the overidentifying restrictions implied by the structure of the later period are rejected when tested against the VAR of the earlier period: $\chi^2(55) = 179(p = 0.00)$.

Contemporaneous Causal Order of the SVAR 1990:01-2007:06

The Causal Graph



The $\mathbf{A}_0$ Matrix

$$
\begin{array}{r}
NBORRES \\ FFR \\ M2 \\ CPI \\ GAP \\ INDPRO \\ R10Y \\ COMP \\ R3M \\ BORRES \\ M1 \\ M2OWN
\end{array}
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \alpha_{65} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \alpha_{76} & 1 & 0 & 0 & 0 & 0 & 0 \\
\alpha_{81} & 0 & 0 & 0 & 0 & 0 & \alpha_{87} & 1 & 0 & 0 & 0 & 0 \\
0 & \alpha_{92} & 0 & \alpha_{94} & 0 & 0 & \alpha_{97} & 0 & 1 & 0 & 0 & 0 \\
\alpha_{101} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\alpha_{111} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & \alpha_{122} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{1211} & 1
\end{bmatrix}
$$

Fig. 3 Contemporaneous causal order of the SVAR 1990:01-2007:06

Figure 4 shows the impulse-response function of *CPI* to *FFR* in each period.[12] Despite the presence of commodity prices and a rich monetary/financial sector, the price puzzle persists in the earlier period. A 1 %-point shock to the Federal funds rate results in a statistically significant increase of a little more than a 1/3 of a percentage point in the CPI at the peak, 13 months after the shock.

While the overall shape of the impulse-response function is similar in the later period, it is statistically insignificant throughout. Our empirical identification is consistent with the work of Hanson (2004) and others who found that including commodity prices as an indicator of future inflation reduced the duration of the positive price response to a positive interest rate shock, but did not completely eliminate it and that the price puzzle is more pronounced before 1982 than after.

[12] Complete impulse-response functions are reported on the Hoover's website: http://public.econ.duke.edu/~kdh9/.

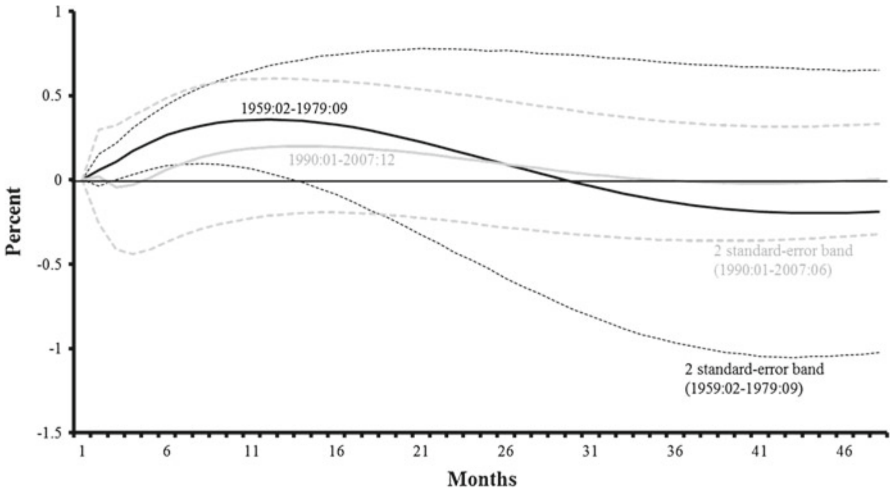


Fig. 4 Response of *CPI* to an *FFR* Impulse (1 % point)

The price puzzle is, then, still a puzzle. Since our empirically based identification potentially incorporates many of the features that various commentators have proposed to resolve the price puzzle, we can still ask what light our SVARs shed on the proposed resolutions. In the earlier period, when the positive response of prices to a Federal funds rate shock persists, do the various solutions offered to the price puzzle—even though they do not solve it—move in the right direction? And equally, though the response of prices is statistically insignificant in the latter period, is that insignificance attributable to the proposed solutions? We consider several of the most prominent proposed solutions.

5.2 Information and marginal cost channels

Sims's (1992) idea that including commodity prices would resolve or reduce the price puzzle by giving policymakers information about expectations of inflation suggests one of two mechanisms. The first mechanism is *direct* and treats commodity prices as a measure of the marginal cost of inputs. The *marginal cost channel* implies a direct causal path ($COMPRICE \rightarrow CPI$) or an indirect path from *COMPRICE* through other variables to *CPI*.

Hanson (2004) sees the small contribution of commodities to total inputs as militating against such a "marginal cost" channel. And our evidence also tells against commodity prices contemporaneously causing consumer prices. For the earlier period, Fig. 2 shows that there is no contemporaneous path connecting *COMPRICE* and *CPI*; indeed *CPI* is exogenous. And for the later period, Fig. 3 shows that *CPI* is again exogenous.

The second mechanism is indirect and treats commodity prices as working through an *information channel* in which they serve as a proxy for various *activity* variables (such as prices or aggregate demand, in contrast to monetary policy variables). Such an information channel requires that directly or indirectly some activity variables serve as a common cause of consumer prices and commodity prices. Since *CPI* is exogenous in both periods, it cannot share a common cause with *COMPRICE*.

Either the marginal cost channel or the information channel can successfully account for a role for commodity prices in mitigating the price puzzle only if policy makes use of the expectations responding either (a) directly to commodity prices ($COMPRICE \rightarrow policy\ variables$) or (b) to the activities about which they are supposed to provide information. Mechanism (b) requires that non-*COMPRICE* activity variables are a common cause of *COMPRICE* and of policy variables (*COMPRICE* ← *non-COMPRICE activity variables* → $policy\ variables$). And, of course, (c) policy variables ultimately must also themselves affect the activity variables (with a lag in a recursive system) in order for the information in commodity prices effectively to moderate consumer prices.[13]

There is little support for these mechanisms in either period. With respect to (a) and (b), Fig. 2 shows that in the earlier period the monetary policy variable *NBORRES* is exogenous and so *it is* not the effect of contemporaneous activity variables. And *FFR* is an effect of *CPI contemporaneously*, though not of *COMPRICE* or any other activity variable. Figure 3 shows that in the later period *COMPRICE* is an indirect effect of variables reflecting the real economy, but is not a contemporaneous cause of any variable, including policy variables. In fact, the policy variables (*FFR* and *NBORRES*) are exogenous.

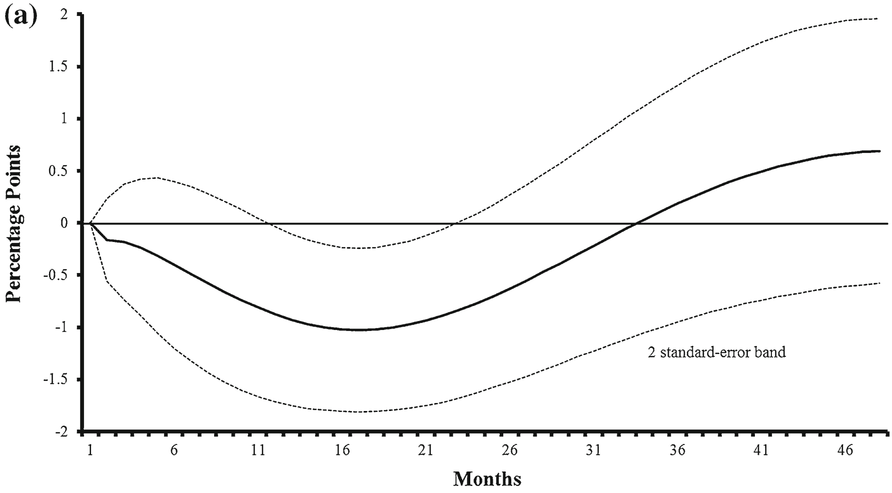
With respect to (c), in the earlier period, the impulse-response functions of industrial production and the output gap to nonborrowed reserves and to the Federal funds rate (Fig. 5) do suggest that the lagged relationship between monetary policy and activity variables is of the right sign. But this is only one part of the needed causal structure. The evidence runs against the other elements ((a) and (b)); and, in any case, the price puzzle is still clearly evident in this period.

The story is not any more favorable to (c) in the later period. The impulse-response functions of industrial production and the output gap to an increase in the Federal funds rate (Fig. 6a, c) are briefly expansionary—the opposite of what one might expect from a contractionary policy. The impulse-responses of industrial production and the output gap to an increase in nonborrowed reserves (Fig. 6b, d) are statistically insignificant and tiny. (These findings are consistent with studies that have found monetary policy shocks having negligible effects on prices—e.g., Leeper et al. 1996, and Christiano et al. 2005.) In all, the two SVARs provide little support for the idea that commodity prices mitigate the price puzzle through an expectations mechanism.

Another way of getting at the role of commodity prices is to consider what happens to the impulse-response functions when their role is suppressed by setting the coefficients on *COMPRICE* at the current and all lagged values to zero. The difference between the actual impulse-response function and this counterfactual impulse-response function measures the work that *COMPRICE* does in the SVAR. If a variable

[13] Many studies have included commodity prices among the activity variables. Some order monetary policy variables after activity variables, assuming that monetary policy affects the activity variables only with a lag (e.g., Strongin 1995; Christiano et al. 1996; Giordani 2004); while others order monetary policy variables ahead of activity variables (Sims 1992; Eichenbaum 1992).

(a)



(b)

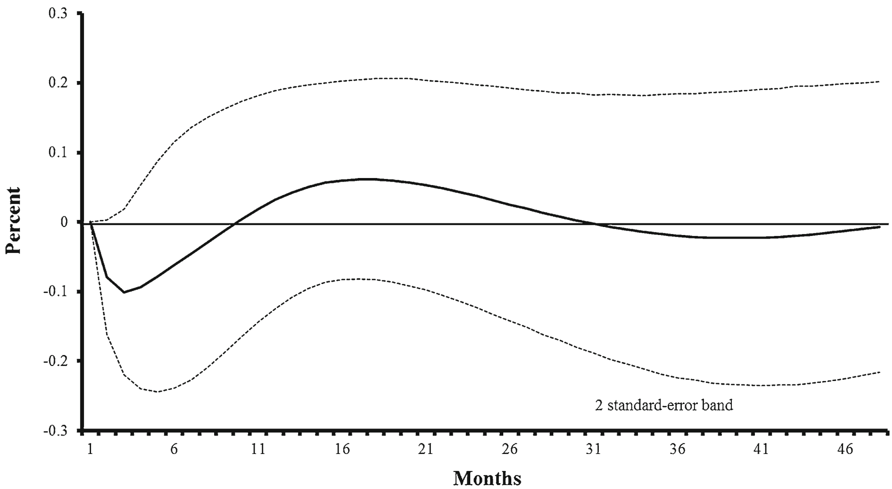
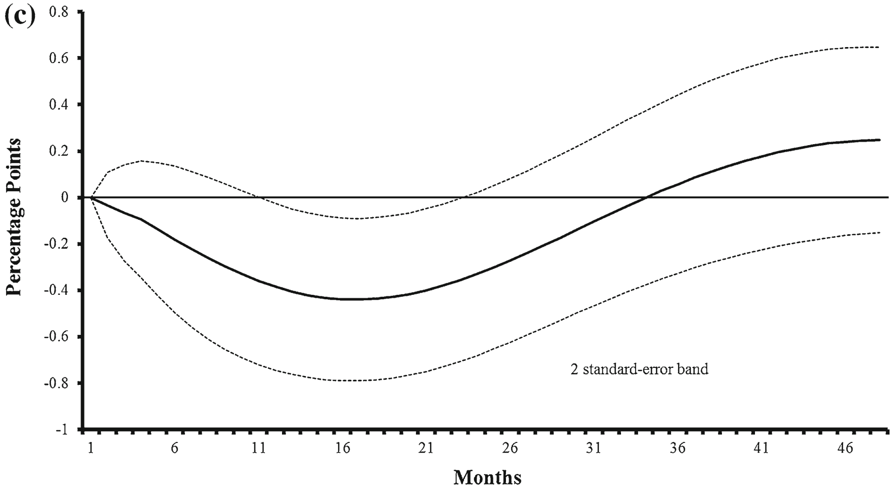


Fig. 5 Impulse responses 1959:02-1979:09. **a** Response of *INDPRO* to an *FFR* Impulse (1 % point) 1959:02-1979:09. **b** Response of *INDPRO* to an *NBORRES* Impulse (1 %) 1959:02-1979:09. **c** Response of *GAP* to an *FFR* Impulse (1 % point) 1959:02-1979:09. **d** Response of *GAP* to an *NBORRES* Impulse (1 %) 1959:02-1979:09

works to mitigate the price puzzle, the impulse-response function with its action suppressed should lie above the original estimate of the response of *CPI* to an *FFR* impulse.

It is important to understand that counterfactuals of this type do not run foul of the Lucas critique. We are not proposing an actual change in policy to which one would have to account for the changes in the decisions of economic actors. Rather we are using the counterfactual merely as a way of quantifying the role of a particular channel

(c)



(d)

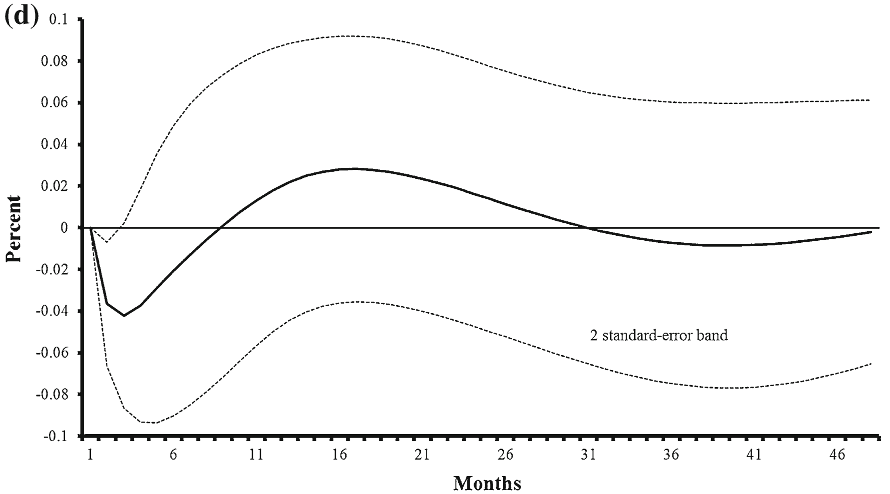


Fig. 5 continued

in a complex web of direct and indirect channels. It is an accounting trick rather than a policy exercise.

Figure 7 shows that *COMPRICE* does not mitigate the price puzzle in the earlier period: the function with *COMPRICE* suppressed is nearly identical to the original function. In contrast, in the later period, Fig. 8 shows that the two functions are substantively and statistically significantly different. The result is, however, odd and offers no support for either of the two expectational mechanisms that we have considered: the function with *COMPRICE* suppressed lies *below* the original estimate. Thus, while *COMPRICE* appears to be an important variable, it does not appear to mitigate the price puzzle, and neither expectational mechanism appears to be compelling.

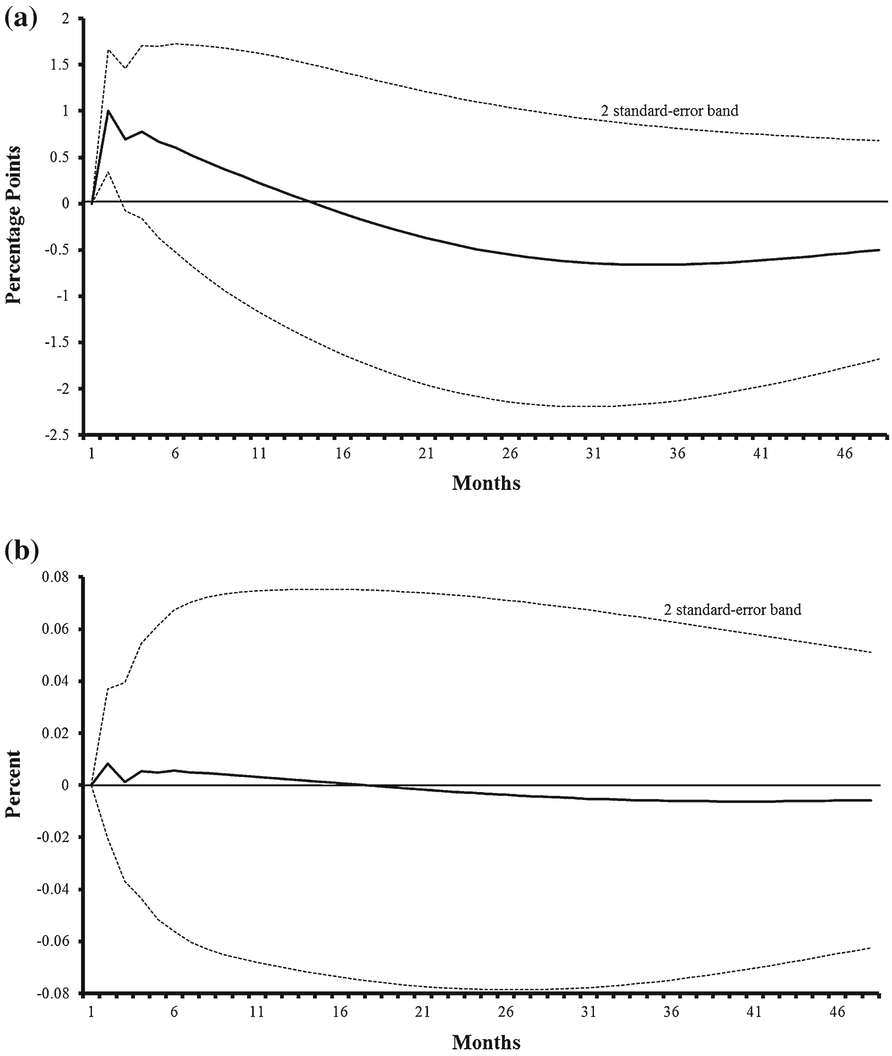


Fig. 6 Impulse responses 1990:01-2007:09. **a** Response of *INDPRO* to an *FFR* Impulse (1 % point) 1990:01-2007:06. **b** Response of *INDPRO* to an *NBORRES* Impulse (1 %) 1990:01-2007:06. **c** Response of *GAP* to an *FFR* Impulse (1 % point) 1990:01-2007:06. **d** Response of *GAP* to an *NBORRES* Impulse (1 %) 1990:01-2007:06

5.3 Omitted measures of real activity

Where many authors looked to the omission of variables that reflect market expectations, Giordani (2004) points to the omission of the output gap from the inflation equation. He speculates that, because interest rates react positively to the output gap, they appear spuriously in the inflation equation with a positive coefficient, and thus act as a proxy for the omitted gap. His explanation requires that the output gap be a

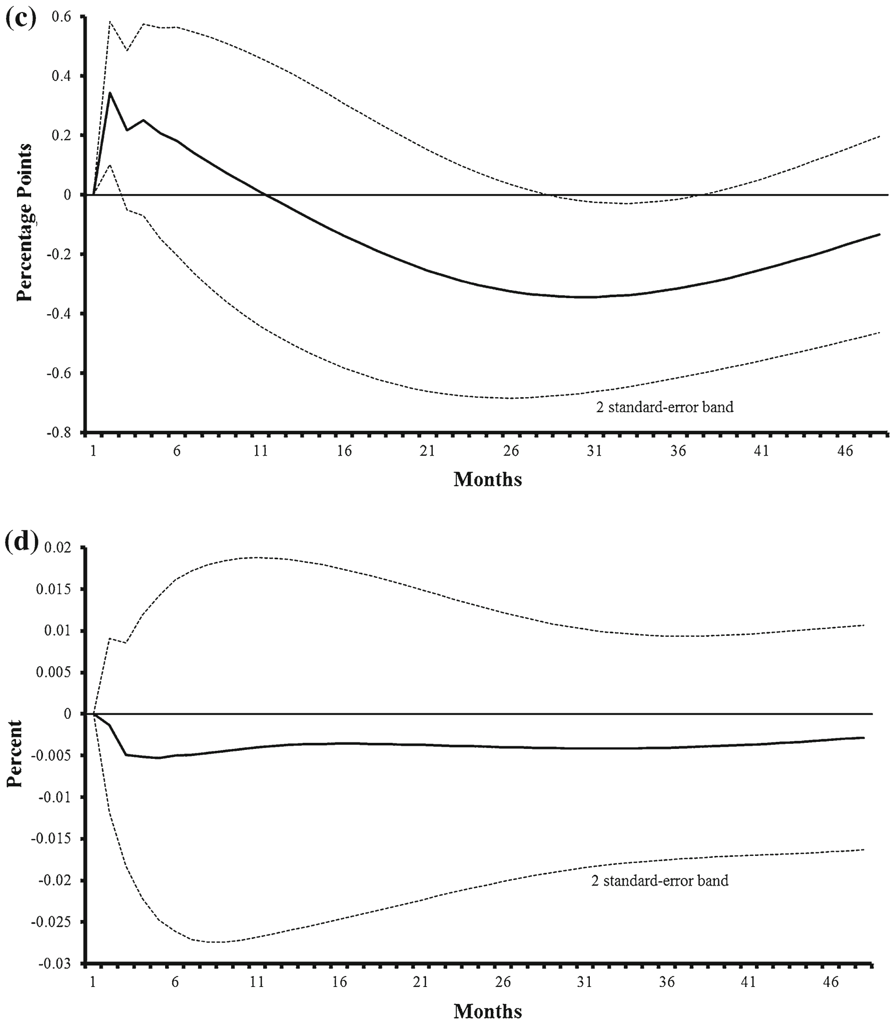


Fig. 6 continued

common cause of interest rates and prices—directly (*interest rates* ← *GAP* → *CPI*) or indirectly. Once the researcher controls for the output gap, the positive relationship between the inflation rate and interest rate should disappear. To the extent that the commodity price index helps to resolve the price puzzle, it is because it is correlated with the output gap. Treated causally this claim requires (again directly or indirectly) *COMPRICE* ← *GAP* or *COMPRICE* → *GAP* or *COMPRICE* ← *other variables* → *GAP*.

In the earlier period (Fig. 2), *GAP* is contemporaneously causally isolated from both prices and interest rates. Do the lagged terms make a difference? Consider the counterfactual response of *CPI* to an impulse to *FFR* when the action of *GAP* is suppressed in all equations. Figure 7 shows that the impulse-response function with *GAP* suppressed

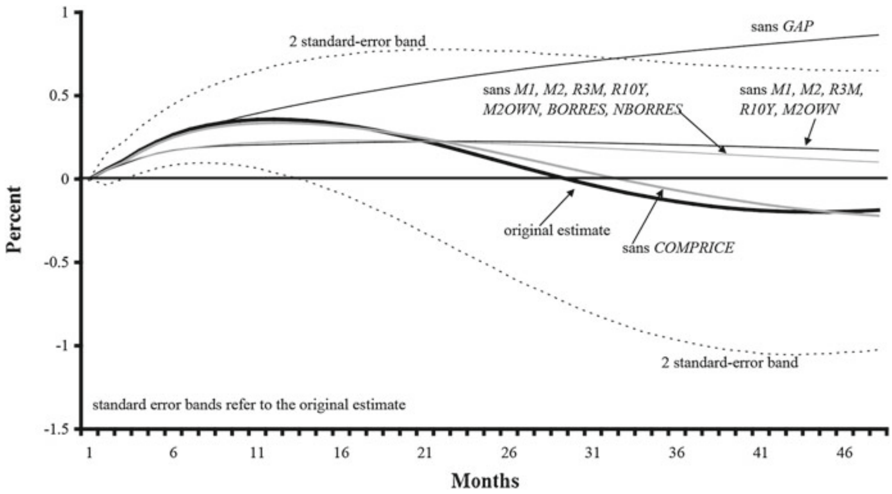


Fig. 7 Counterfactual experiments: response of *CPI* to *FFR* impulse (1 % point), 1959:02-1979:09 with the influences of various variables omitted

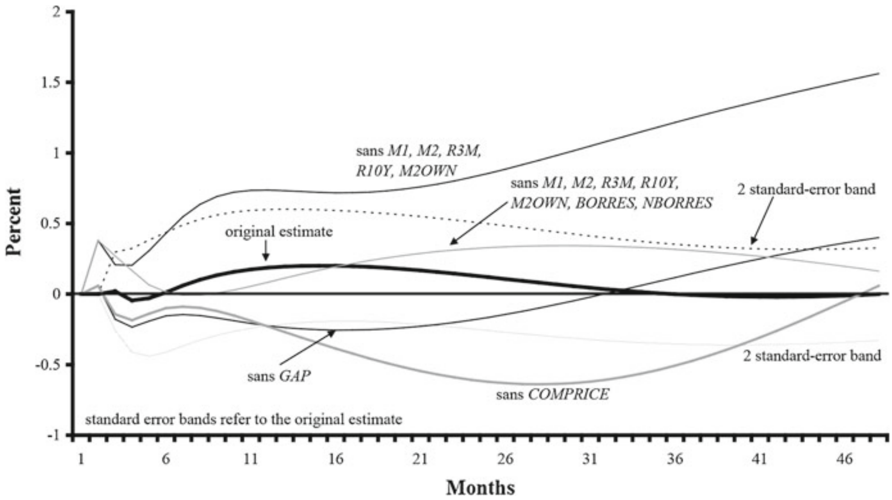


Fig. 8 Counterfactual experiments: response of *CPI* to *FFR* impulse (1 % point), 1990:01-2007:07 with the influences of various variables omitted

lies substantively and statistically significantly above the original impulse-response function. Over the months in which the original function is statistically significant the function omitting *GAP* is nearly indistinguishable; but in later months it continues to rise, suggesting that including *GAP* does in fact mitigate the price puzzle.

Turning our attention to the 1990–2007 period, we must address an ambiguity in the causal ordering. Despite having been oriented according to the dominant direction in the bootstrap replications, the direction of the edge between *GAP* and *INDPRO* is uncertain: *INDPRO* is not an unshielded collider, and a graph with that edge reversed is in the same equivalence class. When this edge is reordered $INDPRO \rightarrow GAP$, the response of *CPI* to an *FFR* impulse (not shown) is virtually indistinguishable from the impulse-response for this period shown in Fig. 4, both in its shape and magnitude and its being clearly insignificant everywhere. Consequently, we continue the analysis with the causal order as presented in Fig. 3.

GAP is contemporaneously causally connected through *INDPRO* to interest rates (*R3M* and *R10Y*) and to *COMPRICE* (Fig. 3). But it is isolated from *CPI*. So, again, the proposed role for *GAP* is not supported contemporaneously. To include lagged effects as well, we again consider the counterfactual response of *CPI* to an impulse to *FFR* when the action of *GAP* is suppressed in all equations. Figure 8 shows that there is a substantive and statistically significant downward shift of the counterfactual impulse-response function compared to the original. *GAP* is clearly important in the later period; but, just as with *COMPRICE*, the impulse-response function shifts in the wrong direction for *GAP* to be a factor resolving the price puzzle.

5.4 The characterization of monetary policy

Another strand in the price-puzzle literature places less emphasis on the role of commodity prices and more on the details of mechanisms of monetary policy (see, for example, Eichenbaum 1992; Strongin 1995; Bernanke and Mihov 1998; Leeper and Roush 2003; Davig et al. 2003). Our SVARs provide a rich characterization of monetary policy and the financial sector in which to investigate its effects.

The price puzzle is usually seen as the perverse response of prices to a shock in the monetary policy instrument. Movements in the Federal funds rate are clearly the principal instrument of monetary policy in the later period. It is by no means clear that the same thing is true in the earlier period. While Bernanke and Blinder (1992) argue that the Federal funds rate is the best measure of the stance of monetary policy in the 1959–1979 period, Federal Open Market Committee (FOMC) directives in the first half of the period typically cite general monetary conditions and the state of reserves—particularly, free reserves—as indicators of the stance of monetary policy, and advocate easing or tightening through open-market operations. The Federal funds rate is mentioned later as one market rate among many, and only near the end of the period are targets set for the Federal funds rate and monetary aggregates. With more targets than instruments and wide target ranges, which are revised at nearly every FOMC meeting, it is hard to see these as binding.

Figure 2 shows that *FFR* is contemporaneously caused by reserve variables (*NBORRES*, *BORRES*), a money market variable (*R3M*), and prices (*CPI*). These might be thought to be the factors considered in the Federal Reserve's reaction function. In that light, the right thought experiment is not to shock *FFR*, holding other factors constant, but to shock monetary policy instruments. The impulse-response functions in Fig. 9a, b show that the Federal funds rate responds in the earlier period in precisely the way that theory suggests: the negative response to a shock to nonborrowed reserves —a shock to the supply side of the reserves market—is consistent with the liquidity effect; while the positive response to a shock to borrowed reserves is what one would expect from a shock to the demand side, though it is borderline significant only for 1 month.

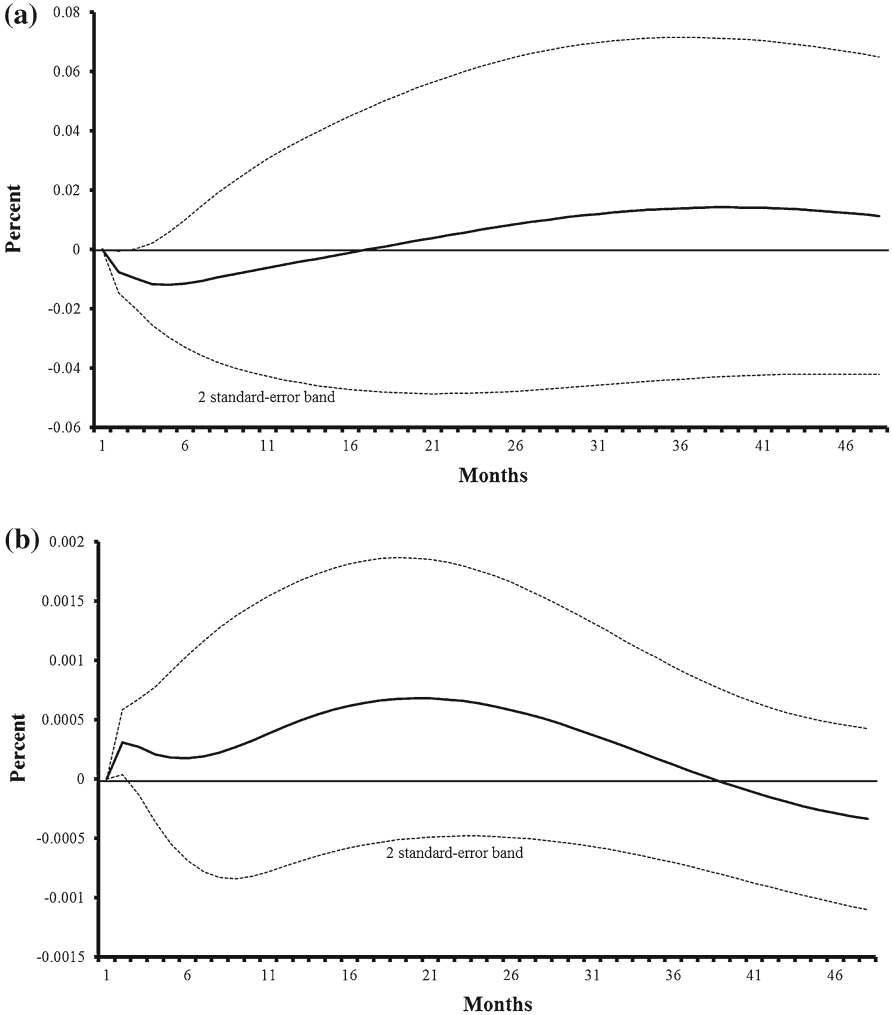
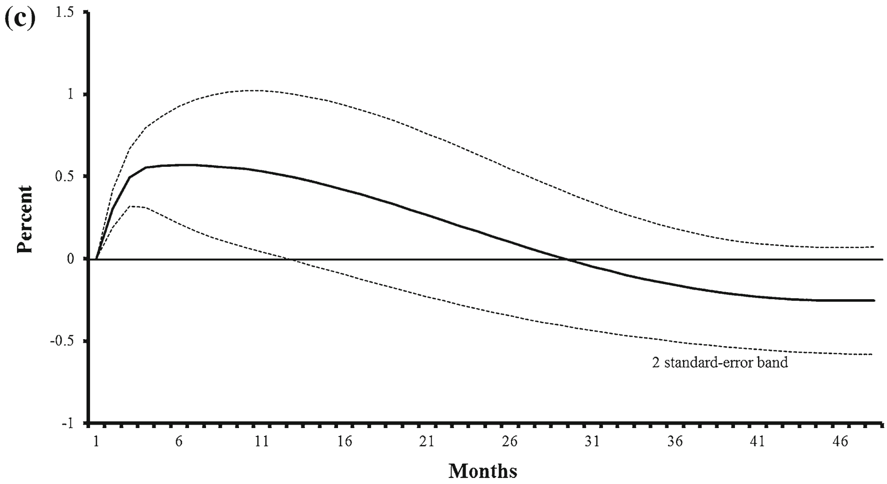


Fig. 9 Impulse responses 1959:02-1979:09. **a** Response of *FFR* to an *NBORRES* Impulse (1 %) 1959:02-1979:09. **b** Response of *FFR* to an *BORRES* Impulse (1 %) 1959:02-1979:09. **c** Response of *FFR* to an *R*3*M* Impulse (1 % point) 1959:02-1979:09. **d** Response of *FFR* to an *CPI* Impulse (1 %) 1959:02-1979:09

This is not, however, the whole story. Shocks to nonborrowed reserves may also capture the Fed's higher frequency accommodation of demand shocks. If the Fed were perfectly accommodating, an independent shock to nonborrowed reserves would not affect total reserves but would be offset by a reduction in borrowed reserves. In fact, the impulse response of borrowed reserves to a shock to nonborrowed reserves (not shown) does offset the nonborrowed reserve shock, albeit incompletely. It would appear that the supply factors dominate demand factors in nonborrowed reserves.

The response of the Federal funds rate to an impulse to the treasury bill rate (Fig. 9c) is positive and significant, consistent with arbitrage in financial markets; while the



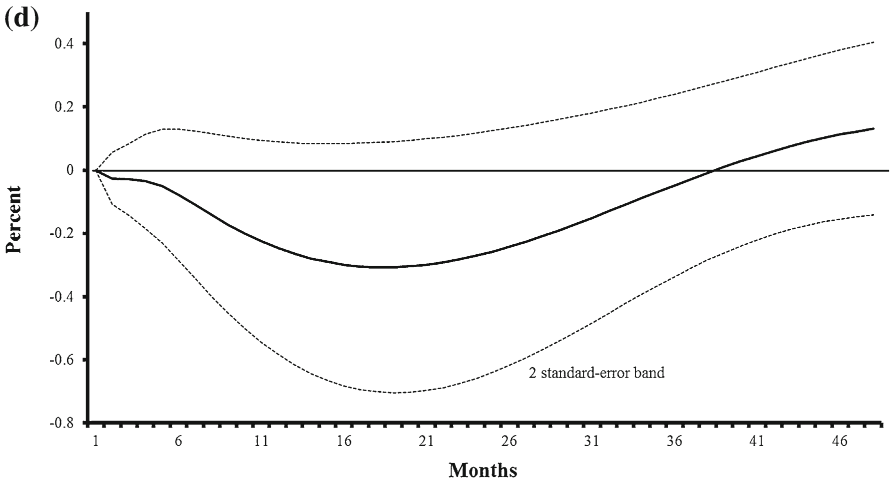
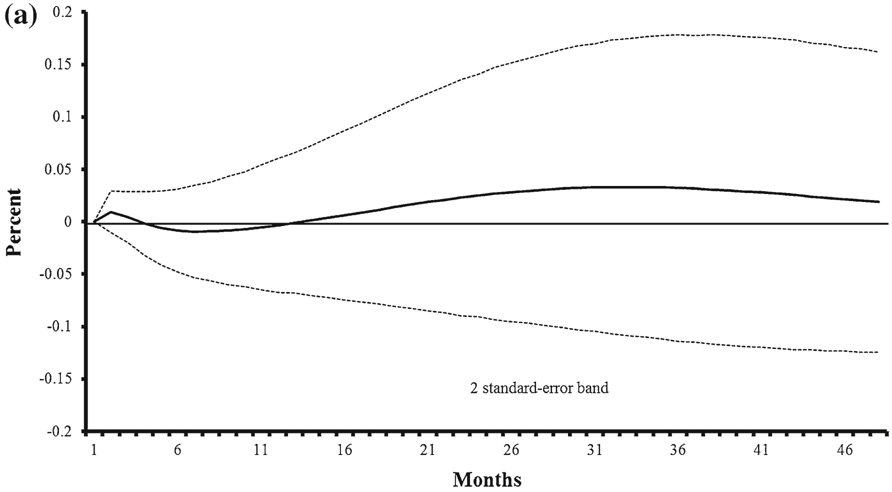


Fig. 9 continued

response to an impulse to the CPI (Fig. 9d) is negative, consistent with a policy of fighting inflation.

The direct effects of the determinants of the Fed's reaction function appear to be sensible, but what about their effects on prices? Figure 10 shows the impulse-response functions of the CPI to each of the reserve variables. Since the Federal funds rate falls in response to an increase in nonborrowed reserves, we would expect theoretically a rise in CPI in response to a shock to *NBORRES*. In fact, the response is negative at first and then positive, but insignificant throughout (Fig. 10a). Likewise, since the Federal funds rate rises in response to a borrowed-reserve shock, we would expect the CPI to fall in response to a shock to *BORRES*. In fact, it does; though once again, the effect is insignificant (Fig. 10b). These results suggest that whether there is a price puzzle in the earlier period, in an important sense, depends critically on whether or not



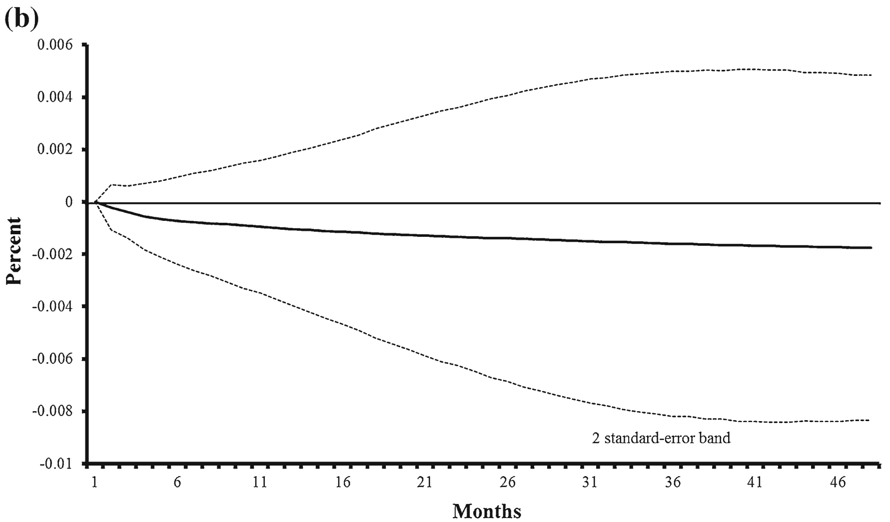


Fig. 10 Impulse responses 1959:02-1979:09. **a** Response of *CPI* to an *NBORRES* impulse (1 %) 1959:02-1979:09. **b** Response of *CPI* to an *BORRES* Impulse (1 %) 1959:02-1979:09. **c** Response of *CPI* to an *R*3*M* Impulse (1 % point) 1959:02-1979:09

monetary policy can be appropriately characterized by actions affecting the Federal funds rate. If the Federal funds rate responds passively to shocks to the demand for reserves and to open-market operations aimed at changing the supply of reserves—a standard characterization of monetary policy in this period—then there is no price puzzle: the perverse price response occurs only through the *unusual* channel of a shock to the Federal funds rate *ceteris paribus* its determinants (particularly the reserve aggregates); whereas an action to tighten monetary policy through the usual channels does not result in a positive price response.

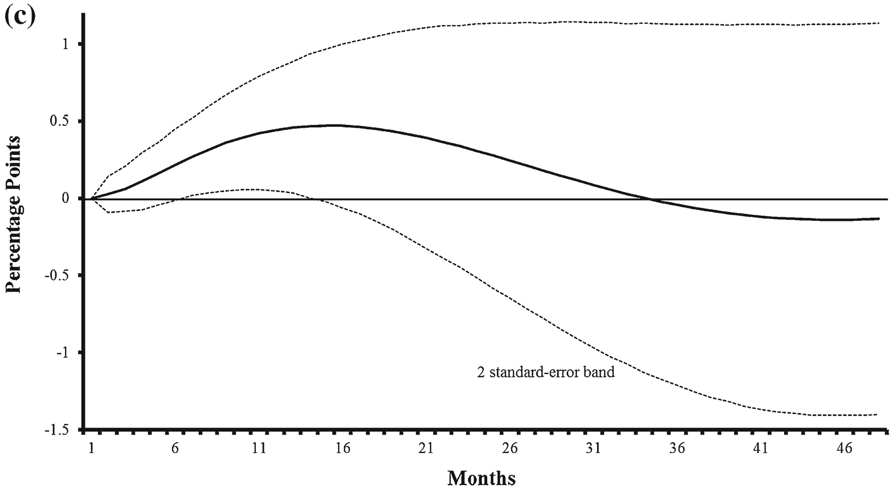


Fig. 10 continued

If the price puzzle is seen a positive response of the CPI to a shock to the Federal funds rate, then, in an important sense, it is not a monetary-policy puzzle at all. This point is driven home by Fig. 10c, which shows the response of the CPI to a shock to another component of the Fed's reaction function—the 3-month Treasury bill. The impulse-response function is nearly identical to that for a direct shock to the Federal funds rate, despite the fact that the initial response of the Federal funds rate to a shock to itself is much greater than to a shock to *R3M*. Taken together, these results suggest that, whatever channel it is that generates the price puzzle and, however, much it may be related to financial markets more broadly, it has little to do with monetary policy per se.

This last conclusion is reinforced by the fact that the impulse-responses of the CPI to the M1 and M2 monetary aggregates (not shown) are positive and significant throughout, while the impulse-response of the monetary aggregates themselves are insignificant with respect to shocks to either reserve measure or the Federal funds rate. The lack of response of the monetary aggregates to the reserve aggregates suggest that textbook money multiplier mechanisms are weak and that M1 and M2 are principally demand determined (Carpenter and Demiralp 2012).

Figure 7 shows two counterfactual experiments in which the effects of (a) the nonpolicy financial system (i.e., *M1*, *M2*, *R3M*, *R10Y*, *M2OWN*) are suppressed in response of *CPI* to an *FFR* impulse and of (b) the reserve aggregates (*BORRES* and *NBORRES*) as well as the nonpolicy financial system are suppressed. The two impulse-response functions are nearly identical. For the early response, if anything, they tend to exacerbate the price puzzle, though for later months they tend to mitigate it. In any case, the two functions lie well inside the ±2 SE bands for the original function. Again highlighting the relative lack of importance of including monetary policy in the 1959–1979 period to resolving the price puzzle.

Figure 8 suggests that things are quite different in the 1990–2007 period. The counterfactual in which the nonpolicy financial system is suppressed lies substantively and significantly above the original estimate of the response of *CPI* to an *FFR* impulse. Accounting for the financial mechanism, therefore, appears to be important in mitigating the price puzzle. A second counterfactual that includes the reserve aggregates complicates matters. Its impulse-response function is substantively and significantly above the original estimate for 2 months (reaching a peak response of about 0.4 points on the *CPI* for a 1 point increase in *FFR*). It then falls reasonably close to the original estimate—and well within the ± 2 SE bands for the original function. While it is reasonable to conclude that the structure of monetary policy and the financial system is important in the later period, exactly how the various pieces interact warrants further investigation.

6 The price puzzle: better understood, if still unresolved

Since the price puzzle was first noticed, it has not been decisively resolved. The various proposed solutions have generally been considered separately and usually in the framework of small SVARs without any common set of variables and typically with ad hoc and casually justified identification schemes. In contrast, we have addressed the price puzzle in a rich 12-variable SVAR. We identify its contemporaneous causal order on the basis of a recently developed graph-theoretic causal search algorithm combined with formal tests of the implied overidentifying restrictions. This SVAR provides a more rigorously justified framework in which to examine the price puzzle and to evaluate the efforts to resolve it.

Our major conclusion is that, even when we move past ad hoc formulations of the SVAR, the transmission mechanism of monetary policy to prices is still puzzling. On the one hand, if we treat the price puzzle as a question of the response of the CPI to monetary policy, then there is no price puzzle evident in either the earlier or the later period. Whereas in the later period, it is reasonable to characterize monetary policy by the behavior of the Federal funds rate, this is less evident in the earlier period, in which policy is more naturally characterized as actions affecting the supply of reserves (such as open-market sales or purchases of nonborrowed reserves) or the demand for reserves (such as changes in the discount rate). But these policy actions do not generate a price puzzle at all.

On the other hand, if we treat the price puzzle, as most of the literature does, as a question of the response of the CPI to shocks to the Federal funds rate, then none of the major resolution-strategies in fact resolves it. On this characterization, a substantively and statistically significant price puzzle persists in the 1959–1979 period. The price puzzle is resolved in the 1990–2007 period only in the sense that there is no statistically significant response of prices to an impulse to the Federal-funds rate, though the impulse-response function is numerically positive. The different impulse responses as well as differences in the causal ordering of the SVARs for each period suggest that there are important structural differences between the two periods. Changes in the structure of monetary policy are likely to be an important element in these structural differences.

None of the common solutions (as identified from the response of prices to shocks to the Federal funds rate) to the price puzzle appear to resolve it. The inclusion of commodity prices in the SVAR in the earlier period makes almost no difference; and, indeed, the SVAR is structured in a way that simply does not support the marginal-cost or informational channels that are offered as a rationale for including commodity prices. Commodity prices do seem to be important for the behavior of consumer prices in the later period, but the way in which they interact does not appear to operate in the direction of mitigating the price puzzle.

In contrast, a measure of the output gap is important in both periods—substantially mitigating the price puzzle in the earlier period, but exacerbating it in the later period. Again, the structural information in the SVAR undercuts the notion that the mechanism of mitigation is related, as has been suggested, to interest rates acting as a proxy for the output gap in specifications that omit the output gap.

While monetary policy does not appear to mitigate the price puzzle in the earlier period, it does appear to mitigate it substantially in the later period. A better understanding the role of the output gap and monetary policy, therefore, should be the focus of future research.

One solution that we did not investigate here is that proposed by Brissimis and Magginas (2006). They suggest that the Conference Board's index of leading economic indicators would be a better proxy for expectations.[14] Unfortunately, this index is not a direct measure of expectations but merely a weighted average of ten series that correlate well with future output.[15] Not only are the component series likely to have a complex causal structure among themselves that would be suppressed in using it as a variable in our SVAR, three of the underlying series, the M2 monetary aggregate, the 10-year T-bill rate, and the Federal funds rate (the last two related in the index as an interest-rate spread) are connected to series in our SVAR as identities, rather than causally. What is more, there is no reason to believe that each of the remaining components of the index stands in the same causal relation to the other variables in the SVAR. Including the index is much more likely to muddle up the actual structure of the economy than to capture some essential feature.

The price puzzle is indeed still a puzzle. The empirically identified SVAR, however, has allowed us to understand more systematically the factors that might matter to the behavior of prices in the face of a Federal-funds-rate shock and to evaluate the mechanisms proposed to account for their actions. This paper will have served its function if it directs research along the more fruitful lines identified.

Acknowledgments Demiralp's research was funded by Turkish Academy of Sciences (TÜBA).

[14] We also do not consider other, and possibly more workable, proxies for expectations such as those based on the Federal funds futures market; but see Faust et al. (2004).

[15] The components of the index are: 1. Average weekly hours, manufacturing; 2. Average weekly initial claims for unemployment insurance; 3. Manufacturers' new orders, consumer goods and materials; 4. Index of supplier deliveries—vendor performance; 5. Manufacturers' new orders, nondefense capital goods; 6. Building permits, new private housing units; 7. Stock prices, 500 common stocks; 8. Money supply, M2; 9. Interest rate spread: 10-year Treasury bonds less federal funds; 10. Index of consumer expectations.

Appendix: specification search

The specification search is documented in Table 3. It proceeds in three stages:

Table 3 Search for contemporaneous causal structure: 1959:02-1979:09

	Specification	Likelihood ratio test against the just-identified model (p value)
Search		
I. Initial model	Graph in Table 1 (causal order selected by the PC algorithm) with undirected and bidirectional edges oriented to the dominant direction under Net Direction	0.15
II. Tests of robustness of imposed directions	As in initial model with edges redirected as:	
A.	*FFR* → *R3M*	0.01
	FFR ↔ *R3M*	0.13*
B.	*M2OWN* → *COMPRICE*	0.16
	M2OWN ↔ *COMPRICE*	0.15*
C.	*R3M* → *M2OWN*	0.00
	R3M ↔ *M2OWN*	Not identified
D.	*FFR* → *CPI*	0.08
	FFR ↔ *CPI*	0.15*
E.	*CPI* → *COMPRICE*	0.15
	CPI ↔ *COMPRICE*	Not identified
F.	Initial model with edges reoriented as starred (*) edges in II(A–D):	
	CPI → *COMPRICE*	Not identified
	CPI ← *COMPRICE*	Not identified
G. Baseline	Initial model with edges reoriented as starred (*) edges in II(A–D) plus *CPI* → *COMPRICE*:	
H.	Baseline and omit *FFR* → *CPI*	0.13
I.	Baseline and omit *M2OWN* → *COMPRICE*	Not identified
J.	Baseline and omit *FFR* *FFR* → *R3M*	Not identified
K.	Baseline and omit *FFR* → *CPI*; and reorient *CPI* ← *COMPRICE*	Not identified
III. Simplification search	Starting with II(H) omit the edges with the lowest t statistics sequentially until all edges are significant at a 5 % critical value. Final specification: Graph in Fig. 2	0.12

I. Estimation of the initial specification with undirected and bidirectional edges set to the dominate direction as shown in the Net Direction column of Table 1: This specification cannot be rejected against a just-identified specification ($p = 0.15$).
II. Test of the robustness of the orientations imposed in stage I: The preferred specification at this stage is II(H).
III. General-to-specific search: Edges with insignificant t−statistics (5 % critical value) are omit *seriatim* until all remaining edges are significant. The final specification cannot be rejected against a just-identified specification ($p = 0.12$).

References

Akleman DG, Bessler DA, Burton DM (1999) Modeling corn exports and exchange rates with directed graphs and statistical loss functions. In: Glymour C, Cooper GF (eds) Computation, causation, and discovery. American Association for Artificial Intelligence and MIT Press, Menlo Park and Cambridge, pp 497–520

Awokuse TO (2005) Export-led growth and the Japanese economy: evidence from VAR and directed acyclical graphs. Appl Econ Lett 12(14):849–858

Barth MJ, Ramey V (2001) The cost channel of monetary transmission. In: Bernanke BS, Rogoff K (eds) NBER macroeconomics annual 2001. MIT Press, Cambridge, pp 199–240

Bernanke BS, Blinder AS (1992) The federal funds rate and the channels of monetary transmission. Am Econ Rev 82(4):901–921

Bernanke BS, Mihov I (1998) The liquidity effect and long-run neutrality. Carnegie-Rochester Conference Series on Public Policy 49. North-Holland, Amsterdam, pp 149–194

Bessler DA, Lee S (2002) Money and prices: U.S. data 1869–1914 (a study with directed graphs). Empir Econ 27:427–446

Bessler DA, Loper N (2001) Economic development: evidence from directed acyclical graphs. Manch Sch 69(4):457–476

Brissimis SN, Magginas NS (2006) Forward-looking information in VAR models and the price puzzle. J Monet Econ 53(6):1225–1234

Carpenter S, Demiralp S (2012) Money reserves, and the transmission of monetary policy: Does the money multiplier exist? J MAcroecon 34:59–75

Christiano LJ, Eichenbaum M, Evans CL (1996) The effects of monetary policy shocks: evidence from the flow of funds. Rev Econ Stat 78(1):16–34

Christiano LJ, Eichenbaum M, Evans CL (1999) Monetary policy shocks: what have we learned and to what end? In: Taylor JB, Woodford M (eds) Handbook of macroeconomics, vol 1A. Elsevier, Amsterdam, pp 65–148

Christiano LJ, Eichenbaum M, Evans CL (2005) Nominal rigidities and the dynamic effects of a shock to monetary policy. J Polit Econ 113(1):1–45

Cooper GF (1999) An overview of the representation and discovery of causal relationships using Bayesian networks. In: Glymour C, Cooper GF (eds) Computation, causation, and discovery. American Association for Artificial Intelligence and MIT Press, Menlo Park and Cambridge, pp 3–64

Davig T, Leeper E, Chung H (2003) Monetary and fiscal policy switching. J Money Credit Bank 39(4): 809–842

Demiralp S, Hoover KD (2003) Searching for the causal structure of a vector autoregression. Oxf Bull Econ Stat 65(Suppl):745–767

Demiralp S, Hoover KD, Perez SJ (2008) A bootstrap method for identifying and evaluating a structural vector autoregression. Oxf Bull Econ Stat 70(4):509–533

Doornik JA, Hendry DF (2007) PcGive Professional 12, vols 1–3 and accompanying PcGive, version 12 software. Timberlake Consultants, London

Eichenbaum M (1992) Comments 'Interpreting the macroeconomic time series facts: the effects of monetary policy'. Eur Econ Rev 36(5):1001–1011

Faust J, Swanson ET, Wright JH (2004) Identifying VARs based on high frequency futures data. J Monet Econ 51(6):1107–1131

Fisher FM (1966) The identification problem in econometrics. McGraw-Hill, New York

Giordani P (2004) An alternative explanation of the price puzzle. J Monet Econ 51(6):1271–1296
Haigh MS, Nomikos NK, Bessler DA (2004) Integration and causality in international freight markets: modeling with error correction and directed acyclical graphs. South Econ J 71(1):145–162
Hamilton JD, Jorda O (2002) A model of the federal funds rate target. J Polit Econ 10(5):1135–1167
Hanson MS (2004) The 'price puzzle' reconsidered. J Monet Econ 51(7):1385–1413
Hendry DF, Krolzig H-M (2001) Automatic econometric model selection using PcGets 1.0 and accompanying PcGets, version 1.0 software. Timberlake Consultants, London
Hendry DF, Lu M, Mizon GE (2009) Model identification and nonunique structure. In: Castle JL, Shephard N (eds) The methodology and practice of econometrics: a Festschrift in honour of David F. Hendry. Oxford University Press, Oxford, pp 343–364
Hoover KD (2005) Automatic inference of contemporaneous causal order of a system of equations. Econom Theory 21(1):69–77
Hoover KD, Perez SJ (1999) Data mining reconsidered: encompassing and the general-to-specific approach to specification search. Econom J 2(2):167–191
Hoover KD, Perez SJ (2004) Truth and robustness in cross country growth regressions. Oxf Bull Econ Stat 66(5):765–798
Hoover KD, Demiralp S, Perez SJ (2009) Empirical identification of the vector autoregression: the causes and effects of US M2. In: Castle JL, Shephard N (eds) The methodology and practice of econometrics: a Festschrift in honour of David F. Hendry. Oxford University Press, Oxford, pp 37–58
Hurvich CM, Tsai C-L (1991) Bias of the corrected AIC criterion for underfitted regression and time series models. Biometrika 86(3):499–509
Juselius K (2006) The cointegrated VAR model: methodology and applications. Oxford University Press, Oxford
Krolzig H-M, Hendry DF (2001) Computer automation of general-to-specific model selection procedures. J Econ Dyn Control 25(6–7):831–866
Leeper EM, Roush JE (2003) Putting M back in monetary policy. J Money Credit Bank 35(6; Part 2): 1217–1256
Leeper EM, Sims CA, Zha T (1996) What does monetary policy do? Brookings Pap Econ Act 2:1–78
Meulendyke A-M (1998) U.S. monetary policy and financial markets. Federal Reserve Bank of New York, New York
Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge
Reale M, Wilson GT (2001) Identification of vector AR models with recursive structural errors using conditional independence graphs. Stat Methods Appl 10(1–3):49–65
Reale M, Wilson GT (2002) The sampling properties of conditional independence graphs for structural vector autoregressions. Biometrica 89:457–461
Scheines R, Spirtes P, Glymour C, Meek C (1994) Tetrad II: users manual. Lawrence Erlbaum Associates, Hillsdale, NJ
Sims CA (1980) Macroeconomic and reality. Econometrica 48(1):1–48
Sims CA (1992) Interpreting the macroeconomic time series facts: the effects of monetary policy. Eur Econ Rev 36(5):975–1000
Spirtes P, Clark Glymour C, Scheines R (2001) Causation, prediction, and search, 2nd edn. MIT Press, Cambridge
Strongin S (1995) The identification of monetary policy disturbances: explaining the liquidity puzzle. J Monet Econ 35(3):463–497
Swanson NR, Granger CWJ (1997) Impulse-response functions based on a causal approach to residual orthogonalization in vector autoregressions. J Am Stat Assoc 92:357–367
Wold H (1954) Causality and econometrics. Econometrica 22(2):162–177
Yang J, Bessler DA (2004) The international price transmission in stock index futures markets. Econ Inq 42(3):370–386