Modelling banknote printing costs: of cohorts, generations, and note-years

Leo Van Hove

Professor of Economics
Department of Applied Economics (APEC)
Vrije Universiteit Brussel (Free University of Brussels)
Pleinlaan 2
B-1050 Brussels, Belgium
+32 2 629 21 25
Leo.Van.Hove@vub.ac.be

Menzies (2004) uses a ‘cohorts approach’ to model banknote printing costs. This paper proposes a ‘generational approach’ that allows for more realistic assumptions concerning currency growth and note replacement. The paper shows that Menzies’ claim that the case for polymer banknotes becomes stronger with higher currency demand is an artefact of his model. In most scenarios, the number of notes in circulation does not affect the relative cost effectiveness of paper and polymer; and when it does, the impact goes in the other direction. A second finding is that the ‘note life over unit-cost rule of thumb’ can by misleading.
Words: 100.

JEL classification: L69; O33; E42

Keywords: Banknotes; Currency; Central banks; Printing costs; Polymer

Modelling banknote printing costs: of cohorts, generations, and note-years

Leo Van Hove

This version: August 28, 2013

Menzies (2004) uses a 'cohorts approach' to model banknote printing costs. This paper proposes a 'generational approach' that allows for more realistic assumptions concerning currency growth and note replacement. The paper shows that Menzies' claim that the case for polymer banknotes becomes stronger with higher currency demand is an artefact of his model. In most scenarios, the number of notes in circulation does not affect the relative cost effectiveness of paper and polymer; and when it does, the impact goes in the other direction. A second finding is that the 'note life over unit-cost rule of thumb' can by misleading.
Words: 100.

JEL classification: L69; O33; E42

Keywords: Banknotes; Currency; Central banks; Printing costs; Polymer

1. Introduction

The number of central banks that have partially or fully adopted polymer banknotes is on the increase. The Bank of Canada (BoC) is a recent high-profile convert. The BoC issued its first 'plastic' note in November 2011 and by the end of 2013 five of its denominations will be printed on the new substrate (Bank of Canada, 2012). The Bank of England (BoE) could follow suit in the next few years. At least that is what Mark Carney, who introduced polymer notes to Canada, suggested soon after his appointment as BoE governor [1].

Apart from improved protection against counterfeiting, for central banks an important advantage of polymer banknotes lies in their higher durability and resulting lower maintenance costs. A downside is that plastic notes are more costly to produce. In a rare academic article on the topic, Menzies (2004) conducts a cost-benefit analysis of the Reserve Bank of Australia's (RBA) migration to polymer, and thus cannot avoid studying the key trade-off between higher durability and higher initial production costs. To that end, Menzies develops a model that compares the net present value (NPV) of the central bank's printing costs in a paper and in a polymer banknote regime. One of the key outcomes of the model is that "robust currency growth strengthens the case for polymer" (o.c., p. 359).

[1] Eccles, L., "UK could introduce new banknotes made of plastic says incoming Bank of England governor", *MailOnline*, February 19, 2013, <http://www.dailymail.co.uk/news/article-2280344/>, last visited on August 23, 2013.

At a time when PayPal predicts not only that by 2016 UK consumers will no longer need cash to go shopping on Britain's high streets but that they will not even need a traditional leather wallet [2], Menzies' insight is alarming for central banks that, like the BoC and the RBA, have opted for polymer, or that, like the BoE, are contemplating doing so. PayPal's belief in mobile payments may well prove too optimistic, but a recent report by the UK Payments Council forecasts that cash payments will fall by around a third between 2012 and 2022, largely driven by increased use of contactless cards and mobile phones [3]. In many a country the use of cash is indeed decreasing, and in some the growth of currency in circulation is already slowing down. Returning to the case of Canada, a recent article in the *Bank of Canada Review* points out that the share of cash in retail payments has decreased continuously over the past 20 years (Arango et al., 2012). In the early 1990s, cash accounted for more than 80 per cent of the volume and about 50 per cent of the value of Canadian point-of-sale transactions. Estimates for 2011 put these shares at below 50 and 20 per cent, respectively (o.c., p. 32). However, in terms of currency circulation, cash has held its ground remarkably well. Arango et al. (o.c., p. 32-33) point out that "the value of bank notes in circulation has risen at an annual rate of about 5 per cent since 2000, virtually the same as the growth in aggregate personal expenditures", a phenomenon which Arango et al. attribute to a rise in the use of cash for non-payment purposes. Still, recent figures on the *number* of banknotes in circulation – which is the more relevant metric when studying production issues – give some cause to worry. Indeed, while over 2005-2009 the number of notes of the five Canadian denominations that are being migrated to polymer increased by an average of 4.01 per cent per year, over 2010-2012 this figure is 3.30 per cent [4]. (In 2011 there was even a standstill, but this might in part be due to the introduction, in November of that year, of the new CAD 20 polymer bill.) The point is that current currency growth in Canada would seem to be lower than when the BoC gave polymer the green light. If this trend continues and if Menzies is correct in arguing that currency growth strengthens the case for polymer, the BoC might thus, in retrospect, have been too keen to adopt polymer.

Fortunately for the BoC, this paper demonstrates that Menzies' finding is, in fact, embedded in his model, which puts paper at a built-in handicap when there is positive currency growth (and vice versa when the demand for currency drops). If the evolution in the number of notes in circulation

---

[2] Bishop, J., "PayPal predicts the future of money", *PayPal blog*, May 7, 2012 <https://www.paypal.fr/blog/PayPal-Predicts-the-Future-of-Money/>, last visited on August 23, 20123.

[3] Payments Council, "New report paints picture of 2022 payments", press release, August 19, 2013 <http://www.paymentscouncil.org.uk/media_centre/press_releases/-/page/2670/>.

[4] Own calculations based on Bank of Canada, *Bank of Canada Banking and Financial Statistics*, March 2013, Table K1 - Bank of Canada note liabilities.

is modelled identically for both technologies, it does *not*, under Menzies' assumptions, affect the ratio of the present values of paper and polymer printing costs. This finding triggers the broader point that Menzies' 'cohorts approach' cannot adequately handle currency growth. I therefore develop a 'generational approach' and use this to gradually relax two of Menzies' key assumptions: I move from a scenario with periodical currency growth to one with yearly growth, and from a situation in which all notes of a given denomination disintegrate all at once - what we will term 'sudden decay' - to a situation in which a constant fraction needs to be replaced each year. I find that as long as paper and polymer are compared over infinite horizons, the number of notes in circulation does not affect their relative cost effectiveness, as intuition would suggest. When the horizon is finite, one has to take into account that a substantial portion of the notes in circulation will not yet have expended their full life. Given that polymer notes last longer, this drives up the cost of polymer more than it does the cost of paper. In such a setting, higher currency growth does affect the relative attractiveness of the two technologies, but the impact in fact goes in the opposite direction compared to what Menzies suggests. A second finding is that in more realistic settings the usefulness of the 'note life over unit-cost rule of thumb' is limited. Third, the proposed generational approach also provides an alternative to Bouhdaoui et al.'s (2013) single-period model. In contrast to what is suggested in Bouhdaoui et al., it is shown that discounting is not neutral, especially when the central bank's planning horizon is finite. Finally, an important preliminary remark is that the present paper only studies scenarios with non-negative currency growth. Indeed, and this is an interesting finding in itself, situations with positive and negative currency growth are not by definition symmetric. However, for the sake of brevity, scenarios with negative currency growth are studied in a companion paper (Van Hove, 2013).

The remainder of the present paper is structured as follows. Section 2 first explains Menzies' approach, shows that his insight concerning currency growth results from a modelling mistake, and analyses a corrected version of his model (in which growth in the demand for paper and plastic currency is 100% identical). Section 3 introduces our generational approach and first applies it to a scenario with yearly instead of periodical currency growth, but with an infinite horizon. Sections 4 and 5 then progressively render more realistic the assumptions concerning, respectively, the time horizon and the note replacement pattern. Finally, Section 6 compares the results across scenarios as well as with the findings of the Bouhdaoui et al. model and Section 7 concludes.

2. Menzies' approach: periodical growth, sudden decay, finite horizon

2.1. The cohorts approach

Key to Menzies' approach is that he reasons in 'note-life periods' instead of years. Specifically, he assumes that all banknotes – both replacement notes as well as additional notes to cover increased demand – are printed in batches at the beginning of each period, and that each note lasts exactly the average note life [5]. In other words, if a given denomination has an average note life, d, of two years, all notes of a given batch disintegrate all at once two years after their date of issuance (hence our use of the term 'sudden decay'); in the meantime the central bank does not incur any replacement costs. The advantage of this set of assumptions is that banknotes circulate in well-defined cohorts [6], and that the NPV of the printing costs can be modelled, as Menzies does, by the sum of a geometric progression.

In particular, Menzies' Eq. (1) represents the NPV, over n periods, of the paper printing costs for a certain denomination growing at a rate of growth g and discounted at an interest rate r:

$$\text{Sum}_{\text{paper}} = a + a\,R + a\,R^2 + \ldots + a\,R^{n-1} = a\,(1 - R^n) / (1 - R), \tag{1}$$

with R, the ratio of the geometric progression, equal to $1 + g - r$ [7]. If one then assumes that polymer notes last T times as long as paper, but cost C times as much to print, the NPV of the polymer printing costs for the same denomination can be written as:

$$\text{Sum}_{\text{polymer}} = Ca + Ca\,R^T + Ca\,\{R^T\}^2 + \ldots + Ca\,\{R^T\}^{(N-1)} = Ca\,(1 - \{R^T\}^N) / (1 - R^T). \tag{2}$$

Note that Eqs. (1) and (2) have a different number of terms. Since polymer notes last T times as long as paper, fewer batches of the former need to be printed over any given time frame. (N is the number of polymer print runs.) Algebraically, in order to obtain (2), every term in (1) is removed except for those T periods apart.

---

[5] An implicit assumption here is that the central bank has perfect foresight. I maintain this assumption throughout the paper.

[6] To be clear: Menzies himself does not use this term.

[7] It is unclear why Menzies refrains from using the standard approach to discounting; that is, setting R equal to $(1 + g) / (1 + r)$. $R = 1 + g - r$ is only correct as a first-order approximation, for very low values of r.

Menzies is then interested in the ratio $\text{Sum}_{\text{paper}}/\text{Sum}_{\text{polymer}}$, which is straightforward to interpret. When the ratio is greater than one, polymer is the cheaper option for the central bank – and *vice versa*. In order to make sure that the paper and polymer technologies are compared over the same time frame, Menzies imposes that $n - 1 = T(N - 1)$ [8], which implies that $N = \{(n - 1)/T\} + 1$.

Given this and given the formulas for the sums in (1) and (2), Menzies obtains:

$$\text{Sum}_{\text{paper}} / \text{Sum}_{\text{polymer}} = (1 - R^T) / \{C\,(1 - R)\}, \tag{3a}$$

and if R is close to one:

$$\text{Sum}_{\text{paper}} / \text{Sum}_{\text{polymer}} \cong T / C. \tag{3b}$$

According to Menzies, "both (3a) and (3b) lead to powerful and intuitive insights" (o.c., p. 358). Menzies lists five such insights. Insight no. 5 is that "robust currency growth strengthens the case for polymer" (o.c., p. 359). This does indeed follow from Eq. (3a): "If currency growth g is large, R will be large and vice versa. […] The limit of (3a) as R approaches infinity (a large positive growth in currency) is infinite, leading to the adoption of a polymer substrate" (ibidem). Another way to see this is by writing (3a) as:

$$\text{Sum}_{\text{paper}} / \text{Sum}_{\text{polymer}} = (1 - R^T) / \{C\,(1 - R)\} = (1 / C)\,(1 + R + R^2 + \ldots + R^{T-1}).$$

Menzies then argues that "there is some cut-off point where future savings on notes are not enough to cover the initial [higher] outlays [on polymer], *because fewer notes are being used*" (ibidem; emphasis added). He explores this in a simulation where $T = 4$, $C = 2$, and $r = 0$, and finds that currency circulation would have to decline by more than 35 percent [9] per year to wipe out the future savings from polymer (o.c., p. 360, Figure 1).

[8] Note that this expression will often fail to yield integer values for n or N. This is because it is incorrect. On p. 358, Menzies points out - and correctly so - that "One polymer print run could replace T paper print runs". However, while $n = TN$, this clearly does not imply that $n - 1 = T(N - 1)$. The most straightforward way to derive Eq. (3a) below is to consider N polymer note-life periods – rather than $N - 1$ as Menzies wants to do (o.c., p. 357) – and the equivalent TN paper note-life periods. Note also that Menzies in the end does obtain an expression for (3a) that is correct – at least in his logic – because he makes a second mistake that cancels out the first. On p. 357, Menzies states that "in the period immediately prior to the Nth polymer print run, the cumulative discounted sum of paper print costs is the sum to $n - 1$ terms of the paper geometric progression". However, when considering $N - 1$ polymer periods, the equivalent number of paper periods is not $n - 1$, but $(N - 1)T$.

[9] On replicating this, the correct figure would rather seem to be 45 percent.

Contrary to what Menzies claims, his finding that the relative cost effectiveness of paper and polymer notes is affected by currency growth is, in fact, not so intuitive, and even intellectually puzzling. If there is a cost saving of *x* dollar cent for every paper banknote that is replaced by a plastic one, it is straightforward that the total absolute cost savings will increase with the number of notes in circulation. However, $Sum_{paper}/Sum_{polymer}$ is a *relative* measure. If polymer outperforms paper by *y* per cent when there are, say, 100,000 notes in circulation, surely polymer will outperform paper by the same *y* per cent when there are twice as many notes in circulation from the very start? [10] And should not the same be true when currency circulation increases along the road? [11] The next subsection shows that this is indeed the case (under Menzies' assumptions), and that Menzies' insight results from a modelling mistake.

2.2. There is currency growth and currency growth

In order to refute Menzies' claim that currency growth strengthens the case for polymer, it is helpful to first introduce some additional, clarifying notation. As mentioned in subsection 2.1, Menzies defines *a* as the printing costs per note-life period and – judging from the example in footnote 4 (on p. 357) – also per denomination. Note also that Menzies applies a *currency* growth rate to these printing costs. In other words, if we define *B* as the number of notes of a given denomination that are in circulation, and *p* as the unit printing cost of said denomination - so that $a = pB$ – then Menzies seems to implicitly assume that *p* is constant over time. This is consistent with his assumption that *T* and *C* are constant. In other words, Menzies assumes that the characteristics of the two printing technologies do not change over time.

There is, however, a problem with *B*. On closer scrutiny, comparing Sum_{paper} as expressed in Eq. (1) with $Sum_{polymer}$ as expressed in Eq. (2) puts paper at a built-in handicap for positive currency growth. The essence is that while the pattern of currency growth may *seem* identical for both types of notes – *g* being the same – in practice, as soon as $g > 0$, there are in fact paper note-life periods during which there are simply more paper than polymer banknotes in circulation, making it only normal that increases in *g* worsen the case for paper.

[10] Note that, in line with this intuition, according to Eq. (3a) the initial level of the printing costs, *a*, does not impact the relative cost effectiveness of the two technologies.
[11] Provided, that is, that the comparison is made over a time frame during which notes of both types have expended their full life (so that differences in durability are captured correctly).

Table 1 underpins this point. Table 1 assumes that paper notes have an average life span of two years (d = 2), and that plastic notes last twice as long (T = 2), so that D = 4. If we consider two polymer note-life periods (N = 2), the equivalent number of paper note-life periods, n, is four. In such a scenario, polymer notes are injected into circulation at the start of years 1 and 5 – as indicated by the arrows in the Table. Paper notes are printed at double this frequency. It is then crucial to realise that, for both types of notes, currency growth takes place from one period to another, not within periods. (All notes needed during a period are printed at the beginning of the period, all notes last exactly the average note life, and in the meantime demand does not increase.) If one adds to this the fact that, because of their lower durability, there are T times more periods for paper than for polymer, it becomes clear that the growth pattern is actually different across the two scenarios.

Table 1 - Banknote circulation (d = 2, T = 2, N = 2)

<table>
<tr><th colspan="2">Polymer note-life periods, N</th><th colspan="4">1</th><th colspan="4">2</th></tr>
<tr><th colspan="2">Paper note-life periods, n</th><th colspan="2">1</th><th colspan="2">2</th><th colspan="2">3</th><th colspan="2">4</th></tr>
<tr><td rowspan="4">POLYMER</td><td>Print runs</td><td colspan="4">↑</td><td colspan="4">↑</td></tr>
<tr><td>Volume of notes in circulation</td><td colspan="4">B</td><td colspan="4">$B(1 + g)^2$</td></tr>
<tr><td>Growth rate</td><td colspan="4">-</td><td colspan="4">g^2</td></tr>
<tr><td>Absolute growth</td><td colspan="4">-</td><td colspan="4">$2Bg + Bg^2 = Bg + Bg(1 + g)$</td></tr>
<tr><td rowspan="4">PAPER</td><td>Print runs</td><td colspan="2">↑</td><td colspan="2">↑</td><td colspan="2">↑</td><td colspan="2">↑</td></tr>
<tr><td>Volume of notes in circulation</td><td colspan="2">B</td><td colspan="2">B(1 + g)</td><td colspan="2">$B(1 + g)^2$</td><td colspan="2">$B(1 + g)^3$</td></tr>
<tr><td>Growth rate</td><td colspan="2">-</td><td colspan="2">g</td><td colspan="2">g</td><td colspan="2">g</td></tr>
<tr><td>Absolute growth</td><td colspan="2">-</td><td colspan="2">Bg</td><td colspan="2">Bg(1 + g)</td><td colspan="2">$Bg(1 + g)^2$</td></tr>
<tr><th colspan="2">Years, t</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr>
</table>

Note: circulation is always at start of period.

Table 1 shows that while the volume of notes in circulation is identical at the start of each new polymer note-life period, this is not true for certain paper periods. In period n = 1 (years 1 and 2), banknote circulation in both scenarios is (obviously) equal to the starting level B. In period n = 3 (years 5 and 6), the volume of circulating notes is also the same in both scenarios, namely $B(1+g)^2$. However, in period n = 2 (years 3 and 4), banknote circulation is $B(1+g)$ in the paper scenario vs. only B in the polymer scenario – which clearly makes for an unfair comparison.

Another way of making the same point: when expressed in paper note-life periods, the volume of paper notes increases continuously, but the volume of polymer notes follows a step function. Obviously, increases in T raise the number of periods during which the volume of polymer notes does not change while the volume of paper notes does. In Menzies' simulation on pp. 359-360, T equals 3, implying that paper is put at an handicap during two out of every three paper note-life periods. When expressed in years instead of note-life periods, both series become step functions - with the number of paper notes changing every d years and the number of polymer notes only every D years. This highlights the fact that Menzies' model cannot adequately handle growth in currency demand. Note that in Menzies' calculations for Australia the life span of the (polymer) A$100 note is no less than 32 years, which implies implausibly long periods with constant demand for that denomination.

2.3. Currency growth is neutral

The above raises the question what a model in which the growth in demand for paper and plastic currency is 100% identical would tell us. If we maintain Menzies' assumption that notes last exactly the average note life, one way to ensure that the paper and polymer technologies are compared fairly is to keep the circulation of paper banknotes steady during each and every polymer note-life period; that is, for T consecutive paper note-life periods at a time. Doing so ensures that the demand for both types of notes follows the *same* step function.

This said, the timing of expenditure will be different. Indeed, while the polymer print runs take place at the start of each period, the corresponding T paper print runs (which are now all of equal size) are spread out - presaging that the discount factor will have an impact. Note that in order to be able to properly take this into account, we need to depart from Menzies' atypical approach to discounting – see footnote 7 – and start working with $1/(1+r)$, as is common practice. Note also that in order to derive the equivalent of Eq. (3a) in the new setting, it is in fact sufficient to consider a single polymer note-life period ($N = 1$) because all polymer periods are identical [12].

Given the above assumptions and with $N = 1$ and $1/(1+r) \neq 1$, Eqs. (2) and (1) become respectively:

[12] The working paper version of the present paper contains a general derivation.

$$\mathrm{Sum}_{\mathrm{polymer}} = Ca. \tag{2'}$$

$$\mathrm{Sum}_{\mathrm{paper}} = \quad a + a\frac{1}{1+r} + \ldots + a\frac{1}{(1+r)^{T-1}} = a\frac{1-\left(\frac{1}{1+r}\right)^T}{1-\frac{1}{1+r}} \quad . \tag{1'}$$

Using (1)' and (2)', a correct comparison of the cumulative printing costs then reads:

$$\frac{Sum_{paper}}{Sum_{polymer}} = \frac{1}{C}\frac{1-\left(\frac{1}{1+r}\right)^T}{1-\frac{1}{1+r}}. \tag{4}$$

This expression is very similar to Menzies' Eq. (3a), but the crucial difference is that there is no sign of g. In other words, we find that – as intuition would suggest – currency growth does not affect the relative cost effectiveness of paper and polymer.

3. The generational approach: yearly growth, sudden decay, infinite horizon

Even an adjusted version of Menzies' approach, as above, only allows currency to grow in steps. How would the analysis change if we allowed for growth to take place every year? The problem with such a setting – and this may well be the reason why Menzies opted for periodical growth – is that notes no longer circulate in well-defined cohorts. Because of this, comparing paper and polymer over any number of years will invariably yield results that are affected by the fact that there are notes in circulation that are not yet worn out – notes of both types, but, crucially, in unequal quantities and with varying remaining note lives [13]. A way around this problem consists in (1) working not with a finite but with an infinite horizon, (2) considering 'generations' of banknotes (notes injected into circulation in a specific year), and (3) analysing the replacement pattern of each individual generation over all future periods. The idea behind using an 'infinite horizon' is to allow all notes to live out their natural life, maintain the assumption (used in Section 2) that N polymer note-life periods equate with TN paper note-life periods, and in this way ensure a correct comparison.

In the cohorts approach, all notes of a given denomination are introduced (and retired) at the same time, and only one cohort per denomination circulates at any one time. In the generational approach, for each denomination several generations of notes – injected in different years and thus

[13] There is one exception, namely when $T = 1$. In such a setting, paper and polymer notes are replaced with the same frequency. But then there is no trade-off between higher durability and higher cost, and hence no need for models like those developed in this paper.

with different remaining note lives – circulate concurrently. In particular, there is an initial, bulk injection at the start of year 1 – a ‘prime’ – corresponding to the total demand for that denomination in year 1 [14]. Notes of this generation are replaced every *d (*paper) or *D* years (polymer). At the start of year 2, provided that demand is higher, the central bank issues a second generation of notes to accommodate the increase in demand - the initial demand already being catered for by the first generation. This injection will give rise to a second note renewal cycle, on top of the first but with a different amplitude and timing. Generation three then adds a third layer, etc. Table 2 visualises this, for the same example as in Table 1. (The relevance of the shaded years and print runs is explained later.)

Table 2 – Timing of print runs: generational approach, sudden decay (*d* = 2, *T* = 2, *N* = 2)

Generations		Years													
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
POLYMER	G1	↑				↑									
	G2		↑				↑								
	G3			↑				↑							
	G4				↑				↑						
	G5					↑				↑					
	G6						↑				↑				
	G7							↑				↑			
	G8								↑				↑		
PAPER	G1	↑		↑		↑		↑							
	G2		↑		↑		↑		↑						
	G3			↑		↑		↑		↑					
	G4				↑		↑		↑		↑				
	G5					↑		↑		↑		↑			
	G6						↑		↑		↑		↑		
	G7							↑		↑		↑		↑	
	G8								↑		↑		↑		↑

It is at this point in the paper that I have to explain why, as mentioned in the Introduction, I will only consider scenarios with non-negative currency growth. The reason is that the approach just explained only works if demand for the denomination under consideration increases (or, in the

[14] Just like Menzies I thus assume a ‘big bang’ introduction of both polymer and paper notes. One way to look at the set-up is to see it as a situation where the central bank has decided to redesign its banknotes (or has to introduce a new currency) and is pondering the merits of polymer and (new) paper.

limit, is constant). If demand drops, the central bank does not need to add layers to the original prime. Rather, the problem is the opposite: there will be redundant notes that end up in the vaults of the central bank (and could be re-injected later, thus reducing the volume of later print runs). In short, the situation is substantially different and I do not have the space to treat it in full in the present paper. The interested reader is referred to Van Hove (2013).

In situations with non-negative yearly currency growth and an infinite planning horizon, it is fairly straightforward to see that $\text{Sum}_{\text{paper}}/\text{Sum}_{\text{polymer}}$ will not depend on the growth rate of currency demand. Let me first stress that both currency growth and the interest rate are now measured on a yearly basis. I will use the symbols *g'* and *r'* to differentiate them from their periodical counterparts earlier in the paper. Let me also make explicit the assumptions that *d, D*, and *T* are all integer (with $D = Td$) - so that the cycles of paper and polymer are in sync and we can reason in full years. This does not lead to a loss of generality; one could just as easily reason in terms of months or even weeks.

To start with polymer, the NPV of the printing costs associated with the first generation (G1) is given by:

$$\text{NPV}_{\text{G1}} = \quad Ca + \frac{Ca}{(1+r')^{D}} + \ldots + \frac{Ca}{(1+r')^{D(N-1)}} + \ldots \tag{5}$$

For G2 and G3, the expressions are respectively:

$$\text{NPV}_{\text{G2}} = \quad \frac{Cag'}{(1+r')} + \frac{Cag'}{(1+r')^{D+1}} + \ldots + \frac{Cag'}{(1+r')^{D(N-1)+1}} + \ldots\text{, and} \tag{6}$$

$$\text{NPV}_{\text{G3}} = \quad \frac{Ca(1+g')}{(1+r')^{2}} + \frac{Ca(1+g')}{(1+r')^{D+2}} + \ldots + \frac{Ca(1+g')}{(1+r')^{D(N-1)+2}} + \ldots. \tag{7}$$

Summing across an infinite number of generations – and exploiting the fact that each generation is, by definition, delayed by one year compared to the previous – gives:

$$\text{Sum}_{\text{polymer}} = \sum_{j=1}^{\infty} \frac{Ca}{(1+r')^{D(j-1)}} \left[1 + \frac{g'}{(1+r')} + \frac{g'(1+g')}{(1+r')^{2}} + \ldots \right]. \tag{8}$$

Similarly, the NPV of the paper printing costs is given by:

$$\text{Sum}_{\text{paper}} = \quad \sum_{i=1}^{\infty} \frac{a}{(1+r')^{d(i-1)}} \left[1 + \frac{g'}{(1+r')} + \frac{g'(1+g')}{(1+r')^{2}} + \ldots \right]. \tag{9}$$

Dividing (9) by (8) yields the following expression for the relative cost effectiveness of paper and polymer:

$$\frac{Sum_{paper}}{Sum_{polymer}} = \frac{1}{C}\frac{\sum_{i=1}^{\infty}\frac{1}{(1+r')^{d(i-1)}}}{\sum_{j=1}^{\infty}\frac{1}{(1+r')^{D(j-1)}}}. \tag{10}$$

In (10), both the numerator and the denominator of the second term on the RHS are infinite geometric series. Moreover, since $r' \geq 0$, the absolute value of the common ratio is smaller than 1 as soon as r' differs from 0. Exploiting this, and substituting $D = Td$, we can rewrite (10) as:

$$\frac{Sum_{paper}}{Sum_{polymer}} = \frac{1}{C}\frac{1-\frac{1}{(1+r')^{Td}}}{1-\frac{1}{(1+r')^{d}}}. \tag{11}$$

Again we find that a nor g' play any role. It is also clear from (11) that an increase in C worsens the competitive position of polymer, and that the opposite is true for an increase in T. Furthermore, a comparison of Eq. (11) with Eq. (4) – which captures the relative cost effectiveness under periodical growth – reveals that the two equations are actually equivalent except for the time frame. As a reminder, for Eq. (4) we have reasoned in paper note-life periods, and there are T paper print runs per polymer note-life period. Here we now reason in years, and one paper note-life period is d years. In other words, the frequency of the print runs is now a factor d higher - which is exactly what Eq. (11) reflects.

It is important to realise that Eq. (11) is only valid when the planning horizon is infinite, and that using it as an approximation for a situation with a finite number of polymer print runs, N, is not advisable. Indeed, if one were to transform Eq. (10) into

$$\frac{Sum_{paper}}{Sum_{polymer}} = \frac{1}{C}\frac{\sum_{i=1}^{TN}\frac{1}{(1+r')^{d(i-1)}}}{\sum_{j=1}^{N}\frac{1}{(1+r')^{D(j-1)}}}, \tag{12}$$

one would again obtain Eq. (11), but this would then be biased against paper. As can best be seen in Eqs. (5) to (7), opting for a finite horizon - in the way just proposed - would imply including in the calculations the same number of print runs (N for polymer, TN for paper) *for each and every generation*. However, while N and TN represent the correct number of print runs for the first D and d generations, respectively, this is not the case for later generations. Indeed, banknotes of the final ND-D, respectively ND-d, generations do not need to be replaced as often as those of the first D (d) generations, and notes of the final D (d) generations do not even need to be replaced at all. In Table 2 we have assumed that $D = 4$ and $d = 2$, and we consider two polymer periods or $ND = 8$

years. As can be seen, $N = 2$ ($TN = 4$) is the correct number of polymer (paper) print runs for the first D (d) generations, but is an overestimate for all remaining generations. All shaded print runs in years 9-12 would be erroneously taken into consideration. And what is especially troublesome is that the problem is bigger for paper than for polymer, as can also easily be formally demonstrated [15].

In short, simply replacing ∞ by N does not allow for a correct cut-off (in the example: at the end of year 8). This said, when pondering the substrate of their banknotes, central banks may well have a fixed time horizon in mind. They might, for example, want to redesign their notes, say, every 10 years. In the next Section we therefore try another approach to come up with an expression that is valid for a finite horizon.

4. Yearly growth, sudden decay, finite horizon

In this Section we would like to determine the relative cost effectiveness of paper and polymer under a finite planning horizon, while maintaining, for now, the 'sudden decay' assumption. As explained, the main challenge will be to make sure that we consider the correct number of print runs for all generations. There is, however, another phenomenon that will prove important, in this Section and the next. The issue at hand was already signalled at the start of Section 3, and I, in fact, in that Section opted for an infinite horizon precisely to circumvent it. The issue is that with a finite horizon there will always be notes that, at cut-off, still have some note life left [16]. Suppose that in Table 2 we were to set the cut-off at the end of year 7. At the end of that year, the polymer notes of G2, for example, could then still last another two years. The paper notes of the same generation, for their part, would still have one year left in them. More generally, the central bank in the final *D-1*, respectively *d-1*, years produces notes of too high durability, and thus in a way overinvests. That is, the printing costs are higher then really needed.

Given the higher durability of polymer, it is intuitively clear that the problem is bigger for polymer than for paper. In my example, four out of the seven polymer generations outlive their paper

[15] See the working paper version of the present paper, available upon request.

[16] Menzies, in his article, can avoid the problem by assuming that "existing paper [i]s allowed to live out its natural life before polymer [i]s introduced" (2004, p. 362). That is, in order to avoid having to destroy paper notes that are still fit for circulation, the central bank waits until the end of a paper note-life period. Such optimal timing is no longer possible under our current assumptions. Once one allows for yearly growth, there will always be paper notes that are not yet fully worn out. This holds *a fortiori* once it is assumed, as in Section 5, that notes do not disintegrate all at once, but at a constant rate.

counterparts. However, a caveat is that in order to correctly capture the magnitude of the effect, it is not sufficient to simply look at the number of notes involved. One also has to take into account how many years of usage the notes still have in them. This is not yet crucial in this Section, but will prove vital in Section 6 when analysing differences in results across scenarios.

I therefore introduce the concept of 'note-years'. The basic idea is to measure the outstanding stock of a given denomination not so much in number of notes but in total number of use years. In this view, when the central bank injects, in year one, *B* polymer notes, it has in fact produced *BD* 'note-years' – *D* being the average lifetime of polymer notes. Each year this stock then falls as notes wear out, until new notes are printed and the stock is replenished. With the note-years concept it is fairly straightforward to demonstrate formally that the problem is bigger for polymer than for paper [17]. It can also be shown that, under the current assumptions, it is in fact impossible to compare paper and polymer without encountering the 'excess note-years problem' [18]. The problem can, however, be limited by not putting the cut-off at an arbitrary number of years but instead making sure that at least for G1 - the dominant generation - the horizon covers *N* polymer and *TN* paper note-life periods, so that the problem does not occur for the notes of this generation. For the example in Table 2, this would imply setting the cut-off at the end of year 4 or year 8.

Given this, what is the best approach to come up with finite-horizon formulas? Provided that one treats G1 separately, it is possible to come up with expressions for $\text{Sum}_{\text{polymer}}$ and $\text{Sum}_{\text{paper}}$ by following the same approach as in Section 3; that is, by looking first at individual generations and subsequently summing across generations. G1 – the prime – needs to be treated separately because it is a generation of a special nature, given the 'big bang' introduction of the new notes [19]. Unfortunately, the resulting expressions - not reported here but available upon request - are not very tractable as they consist of three parts, one for G1, one for generations G2 to GD (or Gd), and yet another for all remaining generations.

There is, however, a way to make G1 more similar to the other generations and in this way compress the expressions (somewhat). The solution consists in splitting up G1 in a 'base' and a growth component, as follows (here for polymer):

$$Ca_t = (1+g')Ca_{t-1} = Ca_{t-1} + g'Ca_{t-1}, \tag{13}$$

[17] See the working paper version of the present paper.

[18] To be clear: there are never any excess *notes* as such. All notes that are printed are effectively needed (to either accommodate the increased demand or guarantee banknote quality).

[19] See footnote 14 on this.

with t the year when the polymer notes are introduced and $a_t = pB$ as above. Basically, I just assume here that currency growth between year *t-1* and year t is the same as in later years.

An obvious implication is that all future generations also need to be 'rebased'. Under this slightly altered approach, the NPV of the polymer printing costs associated with the 'base' volume is given by:

$$\text{NPV}_{\text{base}} = Ca_{t-1}\frac{1-\frac{1}{(1+r')^{DN}}}{1-\frac{1}{(1+r')^{D}}}. \tag{14}$$

The intuition here is that these notes need to be printed N times, and this every D years. For the growth component of G1 and for G2, the expressions are respectively:

$$\text{NPV}_{\text{growth component of G1}} = \frac{Ca_{t-1}g'(1+g')^0}{(1+r')^0}\left[1+\frac{1}{(1+r')^D}+...+\frac{1}{(1+r')^{DN}}\right], \text{ and} \tag{15}$$

$$\text{NPV}_{\text{G2}} = \frac{Ca_{t-1}g'(1+g')^1}{(1+r')^1}\left[1+\frac{1}{(1+r')^D}+...+\frac{1}{(1+r')^{D(N-1)}}\right]. \tag{16}$$

As can be seen, these expressions now have a very similar structure – which is precisely what we wanted to achieve with the rebasing. Summing across N generations, as in Section 3, gives:

$$\text{Sum}_{\text{polymer}} = Ca_{t-1}\frac{1-\frac{1}{(1+r')^{DN}}}{1-\frac{1}{(1+r')^{D}}}+\sum_{j=1}^{N}\sum_{y=(j-1)D+1}^{jD}\left[\frac{Ca_{t-1}g'(1+g')^{y-1}}{(1+r')^{y-1}}\sum_{z=0}^{N-j}\frac{1}{(1+r')^{zD}}\right]. \tag{17}$$

If we exploit the fact that the second factor between the square brackets is a geometric series with ratio $1/(1+r')^D$, (17) can also be written as:

$$\text{Sum}_{\text{polymer}} = Ca_{t-1}\frac{1-\frac{1}{(1+r')^{DN}}}{1-\frac{1}{(1+r')^{D}}}+\sum_{j=1}^{N}\sum_{y=(j-1)D+1}^{jD}\left[\frac{Ca_{t-1}g'(1+g')^{y-1}}{(1+r')^{y-1}}\frac{1-\frac{1}{(1+r')^{D(N-j+1)}}}{1-\frac{1}{(1+r')^{D}}}\right]. \tag{18}$$

Again the expression for the NPV of the paper printing costs is obviously very similar:

$$\text{Sum}_{\text{paper}} = a_{t-1}\frac{1-\frac{1}{(1+r')^{dTN}}}{1-\frac{1}{(1+r')^{d}}}+\sum_{i=1}^{TN}\sum_{y=(i-1)d+1}^{id}\left[\frac{a_{t-1}g'(1+g')^{y-1}}{(1+r')^{y-1}}\frac{1-\frac{1}{(1+r')^{d(TN-i+1)}}}{1-\frac{1}{(1+r')^{d}}}\right]. \tag{19}$$

In short, D becomes d (because of the higher frequency of print runs), N becomes TN (because of the resulting higher total number of print runs), and Ca becomes a (because of the lower level of the printing costs).

Luckily, one does not necessarily need to use Eqs. (18) and (19) in full. Indeed, in order to arrive at the insights that we need, in practice one only needs to consider a single polymer print run since all future periods are repetitions of the first [20]. With N = 1 (and n thus equal to T), Eqs. (18) and (19) become:

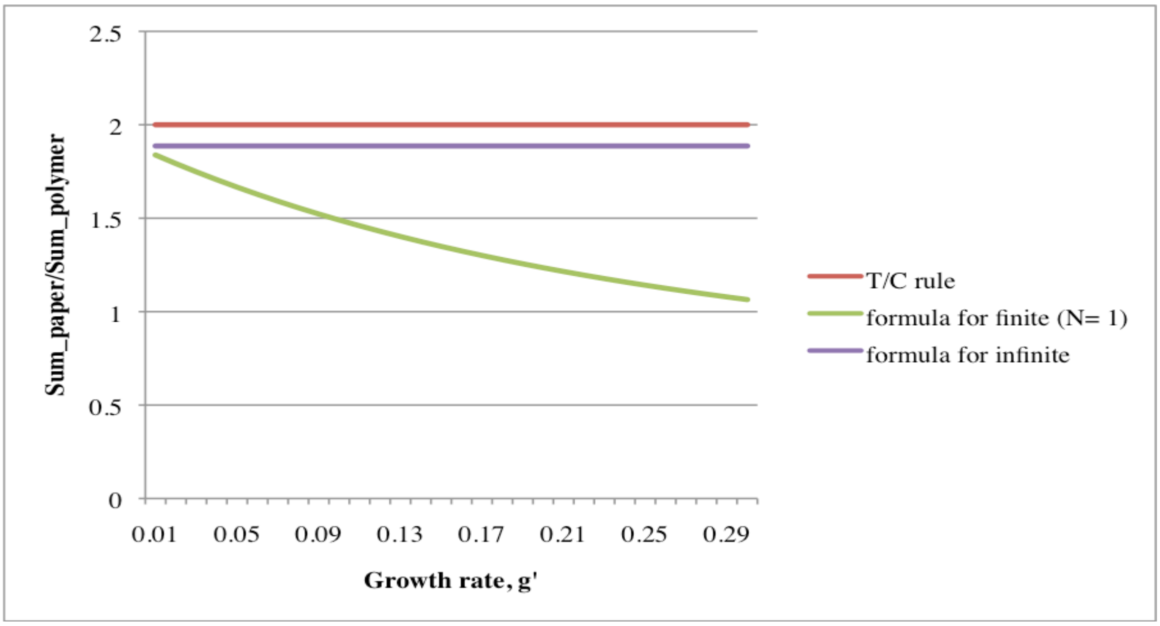
$$\text{Sum}_{\text{polymer}} = Ca_{t-1} + \sum_{y=1}^{D} \frac{Ca_{t-1}g'(1+g')^{y-1}}{(1+r')^{y-1}} = Ca_{t-1} + Ca_{t-1}g' \frac{1-\left(\frac{1+g'}{1+r'}\right)^{D}}{1-\left(\frac{1+g'}{1+r'}\right)}\text{, and} \tag{20}$$

$$\text{Sum}_{\text{paper}} = a_{t-1} \frac{1-\frac{1}{(1+r')^{dT}}}{1-\frac{1}{(1+r')^{d}}} + \sum_{i=1}^{T} \sum_{y=(i-1)d+1}^{id} \left[\frac{a_{t-1}g'(1+g')^{y-1}}{(1+r')^{y-1}} \frac{1-\frac{1}{(1+r')^{d(T-i+1)}}}{1-\frac{1}{(1+r')^{d}}} \right]. \tag{21}$$

Unfortunately, Eq. (21) is still fairly complex. It is, however, quite straightforward to use in simulations of $\text{Sum}_{\text{paper}}/\text{Sum}_{\text{polymer}}$. Figure 2a computes this ratio alternately as (21)/(20) – see the 'finite' series in the graph – and as in Eq. (12) – see the 'infinite' series; the goal being to see how large the pro-polymer bias embedded in (12) actually is; see Section 3. I have also included in the Figure the 'note life over unit-cost rule of thumb' identified by Menzies (2004, p. 359), who states that "Eq. (3b) provides a back-of-the-envelope check on the economic viability of polymer notes. For example, if polymer notes last at least four times as long, and are twice as expensive per note, then (3b) is equal to two and polymer is preferred, because converting to polymer would halve a stream of printing costs". In Figure 1a I have set N = 1, D = 8, T = 4, C = 2, and r'= 0.02. I have set T, the relative durability of polymer vs. paper, at 4 because this is the benchmark value chosen by Bouhdaoui et al. (2012, p. 24-27; 2013) after a round-up of estimates stemming from international experience with polymer notes so far. C = 2 is the figure used by Menzies (2004, p. 362) for the Australian case and is confirmed by Spencer (2011, p. 6) for the case of Canada. By way of sensitivity analysis, in Figures 2b-2d I have, respectively, set T at 3, doubled r' to 0.04, and doubled D to 16 – all the while keeping the other variables at their initial levels.

[20] As I will point out in the next Section, this does not imply that the relative cost effectiveness of paper and polymer is completely identical for N = 1 and N = 2; see footnote 27.

Figure 1a – Relative cost effectiveness: yearly growth, sudden decay, finite horizon



Parameters: $N = 1$, $D = 8$, $T = 4$, $C = 2$, and $r' = 0.02$

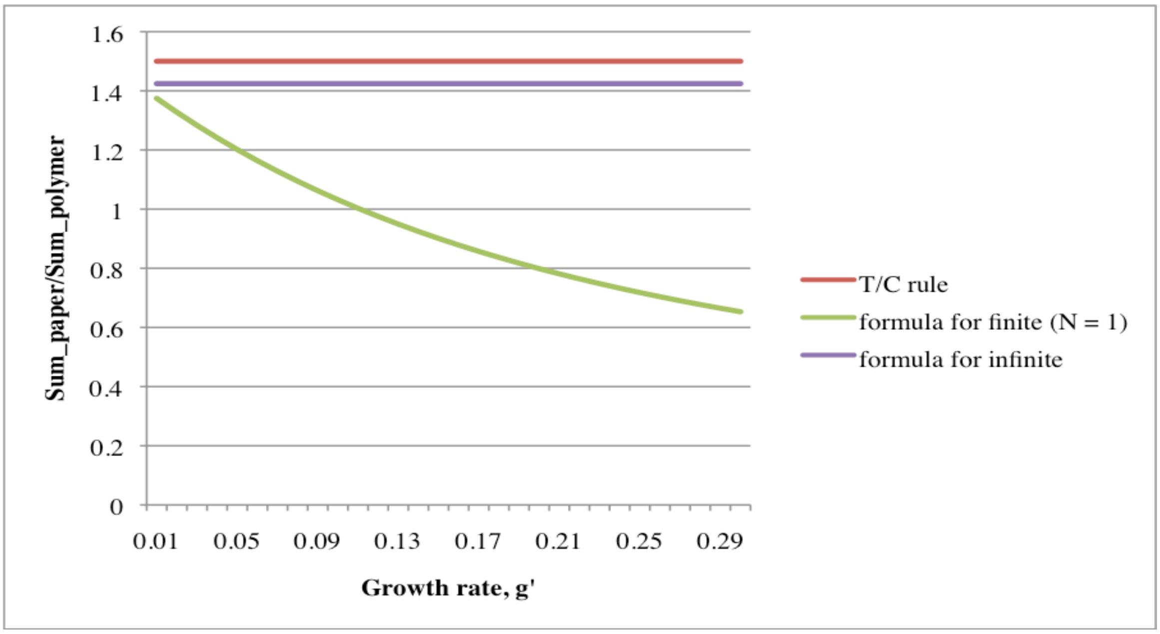
What catches the eye in Figure 1a is obviously that the currency growth rate g' now does impact the relative attractiveness of paper and polymer. By definition this impact cannot come from the first terms in (20) and (21), as these represent the 'base' part of currency circulation. Rather the impact comes from growth in later years in combination with the fixed cut-off point. There are in fact two phenomena at work here. For one, there is the problem detected in Section 3: using our infinite-horizon formula in a finite-horizon setting would overestimate the number of print runs for later generations, and this especially for paper [21]. In other words, the infinite-horizon formula is biased against paper. Our finite-horizon formula was explicitly developed to correct for this. Now, for higher growth rates, the 'weight' of the incorrectly included print runs increases vis-à-vis the earlier runs so that the correction administered by our finite-horizon formula grows in magnitude. Note that in Figure 1a the difference between the infinite and finite series increases with g' [22]. This is, however, also due to a second phenomenon, namely the 'excess note-years issue' discussed above. As explained, the issue occurs in the final years of the planning horizon, and again the relative weight of the print runs, and thus of the problem, increases with g'. This is

[21] This is plainly apparent with the set of parameters used here. With $N = 1$, there is, by definition, only one single print run per polymer generation so that there cannot be any excess print runs. For paper, given that $D = 8$ and $T = 4$, $d = 2$ and $n = 4$. This means that in the infinite-horizon formula there are 4 print runs per paper generation, whereas this is only correct for the first d - that is, two - generations. Generations 3-4 only require 3 print runs, and so on. In short, there is no bias for polymer but there clearly is one for paper.

[22] Note that the two formulas give the same estimate for $g' = 0$ (not visible in Figure 1a). This is only normal. When currency demand is constant, there is only one generation (the prime) so that there are no incorrectly included print runs in the finite-horizon formula, and hence no need for a correction.

bad for polymer because the problem is bigger for polymer than for paper. In other words, there is a negative impact on $Sum_{paper}/Sum_{polymer}$.

Figure 1b – Relative cost effectiveness: yearly growth, sudden decay, finite horizon

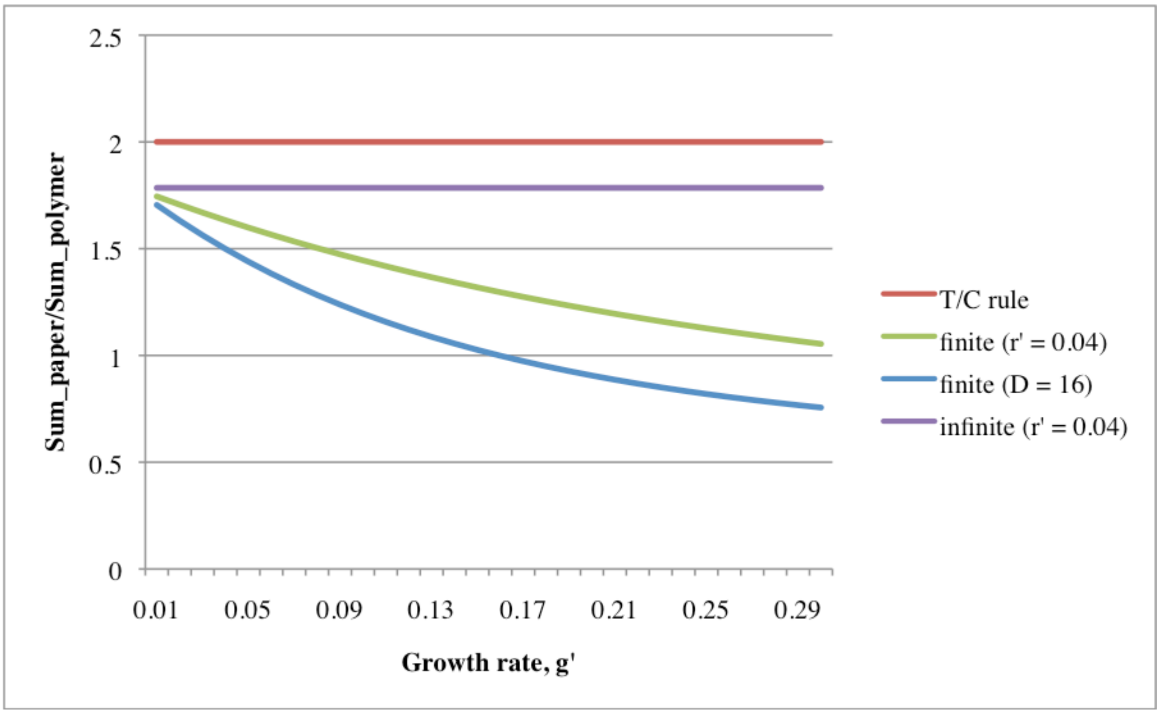


Parameters: $N = 1$, $D = 8$, $\boldsymbol{T = 3}$, $C = 2$, and $r' = 0.02$

A second observation in Figure 1a is that the T/C rule is a poor approximation of the real relative cost effectiveness of paper and polymer, and this even for low values of g'. For $g' = 0.05$, the error already amounts to 20%; the difference between the finite and infinite series being 13%. Figure 1b shows that a lower T – which has now been set at 3 instead of 4 – lowers the attractiveness of polymer, which is intuitively clear. It can also be seen that as g' goes up, the T/C rule continues to indicate that polymer outperforms paper, whereas our finite-horizon formula shows that this is no longer the case for growth rates above 11%.

Finally, Figure 1c shows, first, that a higher discount factor increases the difference between the T/C rule and the infinite-horizon formula (simply because the T/C rule assumes $r' = 0$), but decreases the difference between the finite- and infinite-horizon series. For $g' = 0.05$, the difference is now 11.6% vs. 13.2% earlier. This is because a higher discount factor lowers the weight, vis-à-vis the prime, of both the print runs that are incorrectly included in the infinite-horizon formula and of the excess note-years. At the same time, Figure 1c is another demonstration that the T/C rule does not capture the complete picture. When D is raised from 8 to 16 and T is locked at 4, the finite series shows lower results compared to the benchmark situation in Figure 1a - implying that polymer is less attractive. This is because increasing D, but keeping T constant, increases d (here: from 2 to 4) – as well as the number of years considered (from 8 to 16). Given that for polymer the bulk of the printing costs are incurred upfront [23], whereas for paper the printing costs are more spread out in time, an increase in d benefits paper because the costs are pushed further back in time and will have a lower NPV. This is in line with Menzies' insight no. 1: "If you don't care about the future, use paper" (Menzies, 2004, p. 358) [24].

Figure 1c – Relative cost effectiveness: yearly growth, sudden decay, finite horizon



Parameters of base case: $N = 1$, $D = 8$, $T = 4$, $C = 2$, and $r' = 0.02$

5. Yearly growth, constant disintegration, finite horizon

In this Section we now no longer assume that notes of a given generation disintegrate all at once. Rather we assume that (at the start of) each year a constant fraction of the circulation (of the previous year) needs to be replaced. As symbol for what Menzies (2004, p. 359, footnote 7) calls the 'disintegration rate', I will use $W = 1/D$ (for polymer) and $w = 1/d$ (for paper) – with W and w

---

[23] Remember that $N = 1$, so that there is only one polymer print run per generation.

[24] Another observation in Figure 1c is that the difference between the finite and infinite series is bigger compared to Figure 1a. This is because in a situation where g' exceeds r', (1) the NPV of the incorrectly included print runs increases (so that the correction administered by the finite-horizon formula increases), and (2) the same is true for the NPV of the excess note-years.

referring to the wear and tear that banknotes are subject to [25]. An implication of this change in assumption is that there are no *bona fide* note-life periods anymore and that, apart from the first year (when the circulation is primed), all years are of a similar nature: at the start of the year new notes need to be injected because demand grows at a rate *g'* and replacement notes need to be issued to the tune of $B_{t-1}W$ (for polymer) or $B_{t-1}w$ (for paper). An advantage of this setting is that we no longer need to be fussy when setting the cut-off. Although the timing of the cut-off does affect the relative cost effectiveness of polymer and paper (as I will demonstrate below), the cut-off can essentially be any number of years after launch, say Y – implying that the final print run takes place at the start of that year.

In order, then, to derive new expressions for $\text{Sum}_{\text{paper}}$ and $\text{Sum}_{\text{polymer}}$ we could in principle again rely on our generational approach, make a distinction between initial injections (one per generation) and replacements (*Y-1* times for G1, *Y-2* times for G2, etc.), and then sum across generations. However, in the new setting it is even easier, where replacements are concerned, to simply exploit the fact that each year a fraction W (or w) of the *total* circulation of the denomination under consideration needs to be replaced. $\text{Sum}_{\text{polymer}}$ is then given by:

$$\text{Sum}_{\text{polymer}} = Ca + \sum_{y=2}^{Y} \frac{Cag'(1+g')^{y-2}}{(1+r')^{y-1}} + \sum_{y=2}^{Y} \frac{Cag'(1+g')^{y-2}W}{(1+r')^{y-1}}, \tag{22}$$

with the first two terms capturing the initial injections (in year 1 and the following years, respectively) and the third term capturing the replacements (which obviously only start in year 2). Again we have straightforward geometric progressions and with some re-arranging, Eq. (22) can be written as:

$$\text{Sum}_{\text{polymer}} = Ca + Ca(g'+W)\frac{1}{(1+r')}\left[\frac{1-\left(\frac{1+g'}{1+r'}\right)^{Y-1}}{1-\left(\frac{1+g'}{1+r'}\right)}\right]. \tag{23}$$

Although we reason in terms of printing costs here, the second term in (23) reflects nicely our earlier observation that "demand grows at a rate *g'* and replacement notes need to be issued to the tune of $B_{t-1}W$ (for polymer)"; see the $(g'+W)$ part of the equation. The equivalent expression for paper is completely analogous:

[25] Applying these disintegration rates from the very first year – as I will do – temporarily overestimates the replacement needs. It is only D, respectively d, years after every new injection that the average lifetime will stabilise to D (d). However, this is true for both polymer and paper, and to the same extent.

$$\text{Sum}_{\text{paper}} = a + a(g'+w)\frac{1}{(1+r')}\left[\frac{1-\left(\frac{1+g'}{1+r'}\right)^{Y-1}}{1-\left(\frac{1+g'}{1+r'}\right)}\right]. \tag{24}$$

If the choice the central bank ponders is one between a switch to polymer vs. keeping the existing paper notes rather than between introducing (new) polymer or new paper notes – see footnote 14 on this – Eq. (24) would simply become:

$$\text{Sum}_{\text{paper}} = aw + a(g'+w)\frac{1}{(1+r')}\left[\frac{1-\left(\frac{1+g'}{1+r'}\right)^{Y-1}}{1-\left(\frac{1+g'}{1+r'}\right)}\right]. \tag{24'}$$

That is, in year 1 only a fraction of the notes – corresponding to the 'normal' disintegration – would need to be replaced.

Dividing (24) and (23) shows clearly that the relative printing costs are driven by $1/C$ where the prime is concerned and, apart from the discount factor, by a combination of $1/C$ and $w/W = D/d = T$ in later years; to be more precise, by the factor

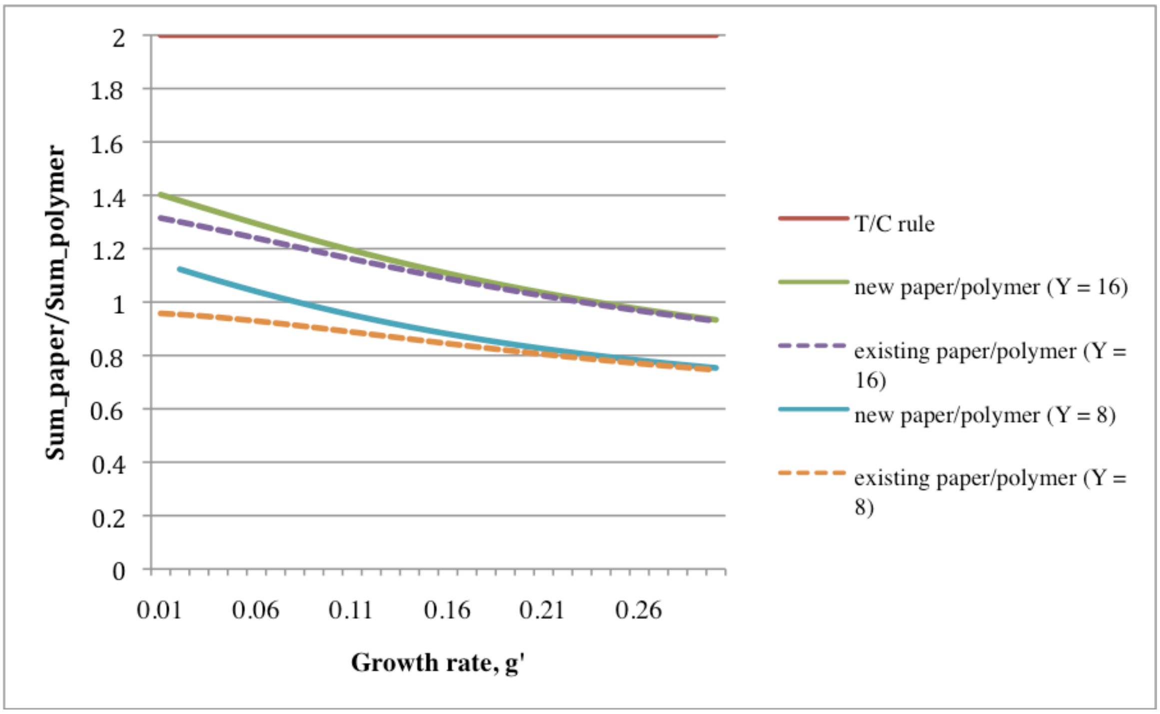
$$\frac{1}{C}\frac{g'+w}{g'+W}. \tag{25}$$

This presages that g' will have an impact on the relative cost effectiveness. Figure 2a analyses this (and other effects) in more detail. Note that all parameters are identical to those in Figure 1a, except for the length of the period considered, which is alternately set at $Y = 8$ (which is equivalent to $N = 1$ in Figure 1a) and $Y = 16$.

A first observation in Figure 2a is that, just as in the previous Section, we again find that higher currency growth negatively impacts the relative attractiveness of polymer. However, this time the impact does not come from a correction for incorrectly included print runs, because there are no such print runs. As explained at the start of this Section, under the assumption of constant disintegration the cut-off itself poses less of a problem. Rather this time the explanation lies solely with the 'excess note-years phenomenon', which also under the current assumptions is bigger for polymer than for paper [26]. When *g'* is raised, the relative magnitude of the problem does not change, but the weight of the generations where the problem occurs does increase vis-à-vis the earlier runs, and this lowers the attractiveness of polymer; see (25).

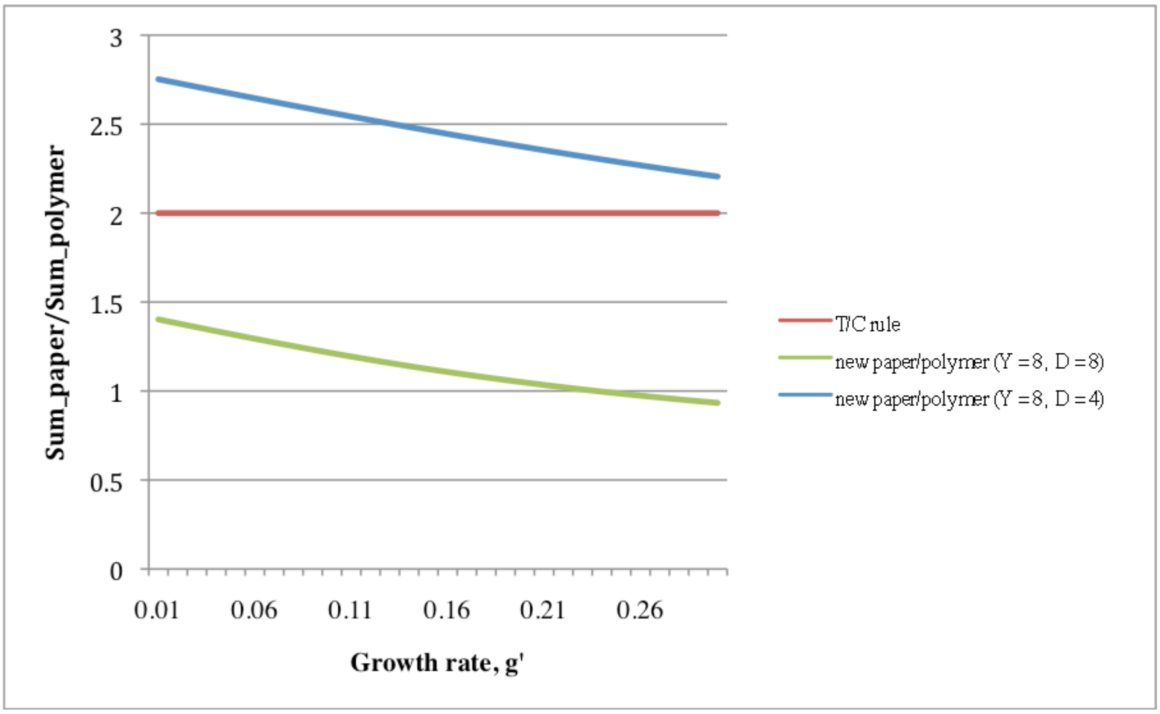
[26] See Appendix 3 of the working paper version of the paper.

Figure 2a – Relative cost effectiveness: yearly growth, constant disintegration, finite horizon



Parameters: $D = 8$, $T = 4$, $C = 2$, and $r' = 0.02$

Figure 2b – Relative cost effectiveness: yearly growth, constant disintegration, finite horizon



Parameters: $Y = 8$, $T = 4$, $C = 2$, and $r' = 0.02$

A second observation in Figure 2a is that an increase in *Y*, all else equal, strengthens the case for polymer. The explanation is straightforward: when the planning horizon is longer, the problem with the excess note-years happens in a more distant future and its NPV is lower [27]. Figure 2b shows another interesting timing phenomenon. In Figure 2b I have lowered *D* from 8 to 4, while maintaining *T* at 4. As can be seen, this dramatically improves the relative attractiveness of polymer. The intuition behind this result is straightforward: the lower *D*, the lower the problem with the excess note-years. Also, in (25), lower values for *D* and *d* imply higher values for *W* and *w*, but with a constant *T*, the factor increases overall, which is bad for paper.

This finding in fact underpins Menzies' insight no. 3: "The benefits of polymer are most dramatic for low denomination notes" (Menzies, 2004, p. 358). Part of the explanation is that high-denomination notes typically last substantially longer than their low-denomination counterparts - for the simple reason that the latter get greater use. For the case of the U.S., Bouhdaoui et al. (2012, p. 23, Table 3) estimate the average life span of a USD 100 bill at 10.4 years vs. a mere 2.7 years for the USD 1 bill. Our results in Figures 2a and 2b show that if the intrinsic costs and benefits of polymer are identical for all denominations (the ratio *T/C* is the same), if all denominations are evaluated over the same horizon (*Y* is the same), and if *g'* and *r'* are the same, then polymer is less attractive for high-value denominations simply because they have a longer life span (a higher *D*). Also, the lower *Y* (the shorter the central bank's planning horizon), the stronger this effect will be. In real life, some central banks effectively use polymer (or a hybrid substrate) for their lower-value denominations but use paper for the higher-value notes. As Kristin Langwasser of the RBNZ pointed out to me in the context of another paper, part of the explanation for such 'mixing' strategy might be that polymer notes can simply be too durable compared to the lifecycle of the note series: "a note lasting 20 years might not be needed if a central bank upgrades and replaces the entire series every ten years" [28].

What is also evident, in both Figure 2a and 2b, is that whereas Menzies (2004, p. 359, footnote 7; my emphasis) claims that "The rule of thumb *also works if we assume* asymptotic depreciation at rate δ, and *a constant fraction of notes disintegrating per period*", the T/C rule in fact again yields poor estimates, for much the same reasons as in Section 4. Returning to Figure 2a, a final, and obvious finding is that if the central bank were to keep the existing paper notes - rather than

[27] Under the 'sudden decay" assumption of Section 4, increasing *N* from 1 to 2 would have a similar impact.
[28] Source: Langwasser, K., personal e-mail, October 11, 2010.

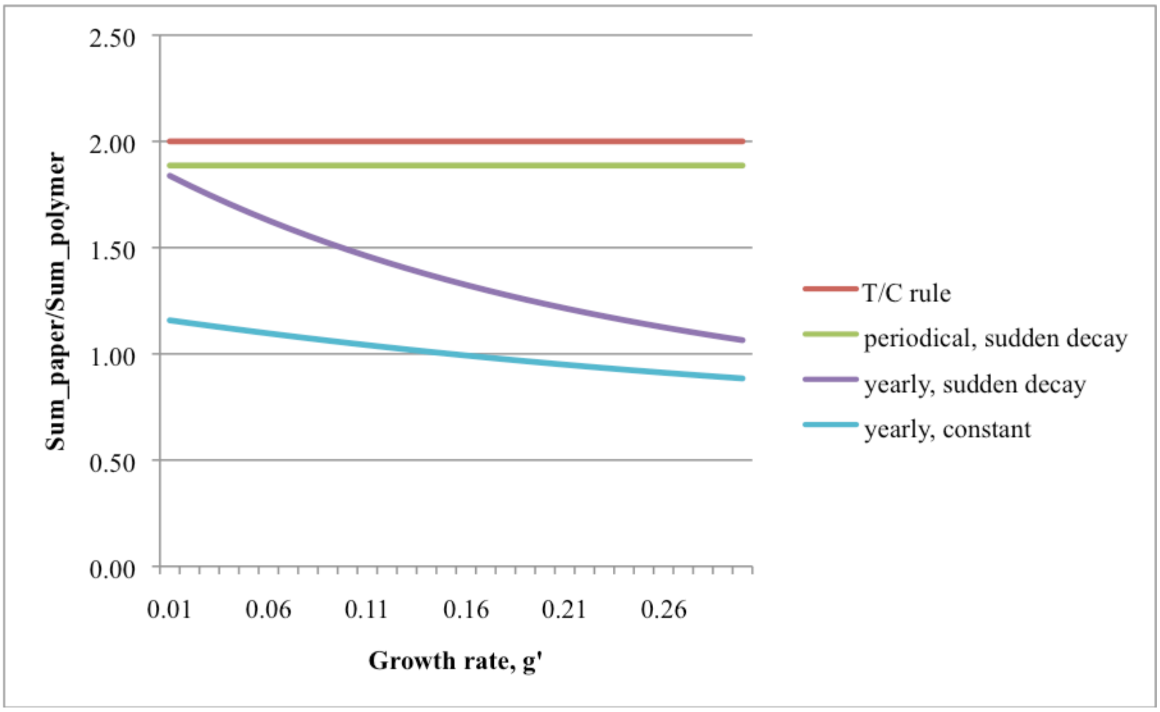
replace them with newly designed paper notes - the case for paper strengthens [29]; compare 'existing paper / polymer' with 'new paper / polymer'.

6. Comparative analysis

Now that we have analysed the NPV of paper and polymer printing costs in a range of scenarios, this Section first compares our results across all scenarios with a finite horizon, and, secondly, with the findings of Bouhdaoui et al.'s (2013) single-period model.

Figure 3 provides an overview of the results for all scenarios with a finite horizon. Starting from the top, the difference between the T/C rule and the other horizontal line - the series for 'periodical growth, sudden decay'; Eq. (4) - is simply due to discounting (which is assumed away in the T/C rule). Higher discounting favours paper because, compared to polymer, more of the expenses are incurred later (in the form of higher replacement costs). In Figure 3 the difference between the two lines is limited, but one has to realise that in our simulation *r'* has been set at a mere 0.02.

Figure 3 – Relative cost effectiveness: comparison across scenarios, finite horizon



Parameters: $N = 1$, D = 8, $T = 4$, $C = 2$, and $r' = 0.02$

[29] Assuming that the notes of the old and new series are equally durable and costly.

Moving downward, the difference between 'periodical growth, sudden decay' (Eq. (6)) and 'yearly growth, sudden decay' (Eqs. (20) and (21)) stems - predictably - from the difference in growth pattern. However, the cause is *not* that in the yearly-growth scenario the frequency of the print runs is higher compared to the periodical-growth scenario and that note issuing thus catches up with demand quicker (so that in certain years the number of notes in circulation is higher). Indeed, this holds true for both paper and polymer, and we are analysing a *relative* measure. As pointed out in Section 3, the *infinite*-horizon formulas for the two scenarios – Eq. (4) for periodical growth and Eq. (11) for yearly growth – are completely equivalent.

The logical inference is that, in a *finite*-horizon setting, the difference between 'periodical growth, sudden decay' and 'yearly growth, sudden decay' comes from … the finite horizon, and in particular from the 'excess note-years phenomenon' that this horizon causes. Indeed, there is no such phenomenon in the periodical-growth scenario whereas it is present in the yearly-growth scenario. As explained in 2.1, under the periodical-growth assumption, banknotes last exactly the full note-life period. Hence, because we always consider N polymer note-life periods and the equivalent TN paper note-life periods, there simply are no excess note-years. Conversely, in the yearly-growth scenario excess note-years are a common phenomenon and the problem is bigger for polymer than for paper, as illustrated in Section 4. Given this, the difference between the lines for 'periodical growth, sudden decay' and for 'yearly growth, sudden decay' grows with g' because a higher currency growth rate increases the weight of the print runs in the years where the excess note-years phenomenon occurs (which are years towards the end of the planning horizon) [30]; see again Section 4.

Finally, if we compare 'yearly growth, sudden decay' and 'yearly growth, constant disintegration' (the two bottom lines in Figure 3), two observations catch the eye: the second line is flatter and lies at a much lower level, especially for lower growth rates. The difference in slope can be explained as follows. As explained in Section 4, with sudden decay, there are *two* reasons why increases in g' negatively impact the case for polymer: a higher growth rate magnifies paper's advantage stemming from the fact that in the final years it requires fewer than the 'normal' TN print runs, and at the same time increases the weight of the print-runs with excess note-years vis-à-vis the earlier runs. As explained in Section 5, with constant disintegration, an 'unfair' difference

[30] While not visible in Figure 3, the two lines intersect for $g' = 0$, which is only normal. When there is no growth in currency demand, there is only one generation of banknotes (namely G1, the prime). As explained in Section 4, for G1 there are no excess note-years because we have made sure that the planning horizon covers an integer number of both polymer and paper note-life periods.

in the number of print-runs is a non-issue so that there is only *one* of the two effects at play. Hence the lower impact of *g'*.

The difference in level, for its part, is due to two reasons. For one, excess note-years are more of a problem with constant disintegration than under sudden decay. With constant disintegration, all generations are affected, even G1. And in our simulations, as long as *g'* is relatively low, G1 is substantially bigger than the other generations. (Notice the difference in distance between the two lines for low, resp. high growth rates). Second, under the assumption of sudden decay and with *N* equal to 1 (as is the case in Figure 3), none of the polymer notes issued in the period considered ever needs to be replaced, as is quite clear from Eq. (20). Conversely, with constant disintegration, as the term indicates, a fraction of all generations, including G1, needs to be replaced every year [31]. This pulls part of the printing costs forward in time, which is an handicap for polymer.

To conclude this Section, it is also instructive to compare our findings with those of Bouhdaoui et al.'s (2012, 2013) model, which is clearly of a different inspiration [32]. Crucially, Bouhdaoui et al. only consider a single (albeit "representative") year. On the other hand, they claim that their model is richer and more flexible compared to Menzies'. For one, Bouhdaoui et al.'s model, next to the cost side, also has a revenue side and as a result can also handle the impact of counterfeiting. Second, whereas Menzies simply reasons in terms of total printing costs per period - and the same is true for the present paper - Bouhdaoui et al. break down these costs into a unit production cost *times* the volume of notes to be printed and also explicitly incorporate processing costs. Third, they claim that their model could be extended to accommodate, for example, a decentralisation of the cash distribution process. There are also marked differences in assumptions, in that Bouhdaoui et al. only consider, in the terminology of the present paper, the scenario with yearly currency growth and constant disintegration.

Despite these different starting points, the 'backbones' of the models by Bouhdaoui et al. on the one hand and Menzies and the present paper on the other exhibit quite a few parallels. First, and unsurprisingly, in Bouhdaoui et al.'s model the core of the analysis also revolves around the crucial trade-off between higher durability and higher production costs, and, although the

[31] Technically, note the presence of the factor *CaW* in Eq. (23), a factor that is absent in Eq. (20).
[32] Note that the 2012 working paper is substantially longer than the published version and contains, for example, a more comprehensive comparison with Menzies (2004).

terminology and the notations differ, the crucial parameters are in fact *T* and *C* [33]. Second, Bouhdaoui et al. also assume a 'big bang' introduction of polymer, and their "one-off migration cost" can be likened to what has been called the "prime" in the present paper. And, interestingly, just as I have often (had) to treat G1 separately from the other generations, Bouhdaoui et al.'s migration cost is in fact outside their main model.

The parallels do not stop here. Some of Bouhdaoui et al.'s findings are also identical to ours (and different from Menzies'). Given the different settings, this does not, however, necessarily imply that they are 100% consistent with ours. For one, Bouhdaoui et al. (2012, p. 18) - who, to remind, consider the case of yearly growth and constant disintegration – argue that "in such a setting, the T/C rule of thumb no longer captures the complete picture". By means of their graphical analysis they show that it is only in the absence of currency growth that the "replacement cost frontier" (their version of the T/C rule) gives the correct answer (o.c., p. 19). Translated to the present paper, the T/C rule would hold thus for $g' = 0$, but only then. This is, however, not fully correct - for two reasons. First, but this is in fact a detail, given their single-year setting Bouhdaoui et al. overlook the impact of discounting. As already demonstrated a number of times, the higher r', the worse the performance of the T/C rule. Second, apart from this point, Bouhdaoui et al.'s inference is only correct for an infinite horizon. As can be seen in Figure 2a, the T/C rule does not hold for finite horizons, not even when $g' = r' = 0$ [34].

A second finding of Bouhdaoui et al. that is – at least at first sight – consistent with ours is their observation that "currency growth would actually seem to *weaken* the case for polymer" (2012, p. 19; emphasis in original). They elaborate, however, that this is purely because their model reasons in terms of *yearly* operating gains for the central bank: "if [growth] is high, then the year in question is a particularly 'bad' year for the central bank, in the sense that it needs to print more additional banknotes, at a higher cost compared to paper banknotes. Hence, the net benefits [of polymer] become smaller and may turn negative in that particular year" (o.c., p. 20; 2013, p. 53). They further state that while high currency growth following the migration to polymer will lengthen the technology's break-even horizon, "given enough time, and provided that [*T*] > [*C*], the central bank will eventually also recoup this additional investment in new polymer notes"

[33] Bouhdaoui et al. (2013, p. 51) talk about "relative cost-effectiveness", $r_p(j)$, and "relative improvement in lifespan", $r_d(j)$. Bouhdaoui et al. (2012, p. 18) explicitly state that their replacement cost frontier "corresponds to Menzies' T/C rule, with a small twist: because we have $r_p(j)$ (Menzies' C) on the vertical axis and $r_d(j)$ (Menzies' T) on the horizontal axis, the frontier should in fact be read as C/T".

[34] This can also be seen by setting g' and r' at zero in Eqs. (23) and (24).

(ibidem). In other words, Bouhdaoui et al. clearly remain convinced that currency growth in fact does not impact the relative cost effectiveness of polymer and paper [35]. But this is because they seem to implicitly reason in terms of an infinite horizon; cf. "given enough time" and "eventually". That a finite horizon can make a difference, just as with the T/C rule above, is something that eludes their static model.

Perhaps our comparison with Bouhdaoui et al. is best rounded off by commenting on this quote of theirs: "Unlike Menzies (2004), our model is single-year instead of multi-period. This is, however, not a weakness. Our model could easily be made multi-period, but the value added would be limited, as each year is simply a repetition of the previous" (2013, p. 58). If anything, the present paper shows that a multi-period model does have its merits: given the difference in the timing of expenses between polymer and paper, discounting is *not* neutral; and the impact of a change in planning horizon cannot be analysed in a single-year model.

7. Conclusions

I have three sets of conclusions. From a modelling perspective, the main observation is that once the assumption of periodical growth is dropped, the timing of the cut-off (when all notes are recalled at once, in a reverse 'big bang') is critical. One option is to make sure that notes of both types can live out their full life – what I have called the 'infinite horizon' – but then the two technologies are in fact compared over different periods. The alternative is to opt for a fixed period – the 'finite horizon' – and then the excess note-years phenomenon is a major driver of the results.

In terms of theoretical results, the main finding is that the intuition formulated at the end of Section 2.1 – 'if polymer outperforms paper for a given level of currency demand, should not the same be true for higher/growing levels?' – is correct as long as one compares N polymer note-life periods with the equivalent number of paper periods, TN. Specifically, the paper first shows that Menzies' (2004) model cannot handle currency growth in a realistic way and, worse, puts paper at an handicap vis-à-vis polymer for positive currency growth. In certain years there are simply more paper than polymer banknotes in circulation, which evidently entails higher printing costs for

[35] In the article version of their paper they are very outspoken about this: "Note that this should not be interpreted as if currency growth weakens the case for polymer" (Bouhdaoui et al., 2013, p. 53).

paper [36]. I adjust Menzies' model and show that when the (periodical) growth pattern is identical for both technologies, the number of notes in circulation does *not* affect the relative cost gap. The generational approach that is developed as an alternative to Menzies' cohorts approach then, step by step, shows that this remains true in scenarios with yearly currency growth and with constant disintegration of notes, at least as long as paper and polymer are compared over infinite horizons. When the horizon is finite, one has to take into account that a substantial portion of the notes in circulation will not yet have expended their full life. Given that polymer notes last longer, in such a setting higher currency growth actually *lowers* the attractiveness of polymer. A second, and related result is that the 'note life over unit-cost rule of thumb' is correct for infinite horizons (and for low discount factors [37]) – see Eqs. (4) and (11) – but only then. The rule overlooks the effects associated with finite horizons. The same can be said, and this is our third finding, about Bouhdaoui et al.'s (2012, 2013) single-period model. Also, in contrast to what is suggested by Bouhdaoui et al., it is shown that discounting is *not* neutral. Discounting in fact has a more ambiguous impact than even Menzies' multi-period model conveys. Given that, compared to polymer, the paper printing costs are spread out more in time, intuition would suggest that a higher r' is bad for polymer. This is true as long as the planning horizon is infinite. But for finite horizons there is a second effect that goes in the opposite direction: a higher discount factor lowers the weight, vis-à-vis the prime, of the excess note-years and – under the 'sudden decay' assumption – also shrinks paper's advantage stemming from the fact that in the final years it requires fewer than the 'normal' TN print runs. While the second effect never overpowers the first, it does imply that higher discount factors are less negative for polymer.

Where policy implications are concerned, a first conclusion is that central banks that need to evaluate the merits of polymer should not be worried about lower currency growth rates, at least not as long as they remain positive; see, however, Van Hove (2013) for the case of negative growth rates. A second conclusion is that, for finite horizons, the T/C rule of thumb can be substantially too optimistic. This is especially true when the planning horizon is short (Y is low), when the denomination has a long life span (D is high), and/or when there is substantial growth in

[36] To be clear: one should not confuse the volume of notes in circulation at a certain point in time (which is a stock variable) with the volume of notes issued over a certain period (a flow variable). Over any polymer note-life period, the *flow* will obviously be larger for paper than for polymer. This is the corollary of the shorter life span of paper notes: they have to be replaced, whereas the initial batch of plastic notes lasts the entire period. My point is about the *stock* of notes. My point is that in Menzies' model, the growth pattern of currency demand is different in the paper and plastic scenario, and that in several years the demand for paper notes is greater than the demand for plastic notes; see my Table 1, where "volume of banknotes in circulation" refers to the stock, *not* the flow.

[37] Note that the T/C rule does not require, as Menzies suggests, that R is close to one (with $R = 1 + g - r$), but that r' is close to zero.

demand (*g'* is high). This said, the central bank can lessen, and *in extremis* even avoid, the excess note-years problem that lies at the heart of these findings by simply not opting for a 'sudden stop'; that is, by, at the end of the horizon, only replacing with new-design notes those polymer notes that are fully worn out and letting the other notes live out (the bulk of) their note life [38]. A third important finding is that the attractiveness of polymer differs between denominations. Under both sudden decay and constant disintegration, I find that polymer is less attractive for high-value denominations. This is because these denominations typically have a longer life span. A higher *D*, ceteris paribus, exacerbates the excess note-years phenomenon, and, in the case of sudden decay, also has the effect of pushing paper print costs further back in time (see Section 4, analysis of Figure 1c). Depending on the life spans of their higher-value denominations and on the planning horizon used, central banks might thus want to consider a partial migration to polymer. This is in line with the findings of Bouhdaoui et al. (2013, p. 57, Table 4), who, for the case of the U.S., find unworkably long break-even horizons for the USD 50 and USD 100 bills. Concerning the latter note they also point out: "The horizon of more than 8000 years signals that if the growth in demand for this denomination remains at the level that we observe in our five-year period (on average 4.8% per year), the cumulative benefits will in practice never exceed the cumulative costs. Indeed, because of the high growth in demand, every year the Fed needs to inject a substantial volume of additional $100 bills, at a higher cost compared to paper, whereas the lower replacement costs take time to materialize" (o.c., p. 57-58). This is an excellent illustration of the fact that growth in demand also plays an important role, as I also find in the present paper.

Finally, there is the unanticipated finding – a finding that has modelling, theoretical, as well as policy implications – that once the periodical-growth environment is abandoned, situations with positive and negative currency growth are no longer symmetric. Given that central banks will ultimately, and perhaps in a not so distant point in the future, be confronted with a less-cash society, this is the scenario that I will analyse in a future paper (Van Hove, 2013).

[38] A drawback is the lengthening of the period during which notes of different designs (and technologies) co-exist.

Acknowledgements

I am indebted to Marc Jegers, Yassine Bouhdaoui, David Bounie, and Steven Vanduffel for comments on an earlier version of this paper. Special thanks go to Gordon Menzies for the open-minded discussion and for pointing out a mistake in one of my early formulas.

References

Arango, C., Huynh, K. P., Fung, B. and **G. Stuber**, The changing landscape for retail payments in Canada and the implications for the demand for cash, *Bank of Canada Review*, Autumn 2012, p. 31-40.

Bank of Canada, Bank of Canada issues $50 polymer bank note, press release, March 26, 2012 <http://www.bankofcanada.ca/wp-content/uploads/2012/03/pr260312.pdf>.

Bouhdaoui Y., Bounie D. and **L. Van Hove**, Production costs, seigniorage and counterfeiting: central banks' incentives for improving their banknote technology, working paper, March 13, 2012 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1652173>.

Bouhdaoui Y., Bounie D. and **L. Van Hove**, When do plastic bills lower the bill for the central bank? A model and estimates for the U.S., *Journal of Policy Modeling*, Vol. 35, Nr. 1, 2013, p. 45-60.

Menzies, G., Money to burn, or melt? A cost-benefit analysis of Australian polymer banknotes. *North American Journal of Economics and Finance*, Vol. 15, Nr. 3, 2004, p. 355-368.

Spencer, C., "Paying with polymer: developing Canada's new bank notes," *Bank of Canada Review*, Supplement, June 20, 2011, p. 1-7.

Van Hove, L., Modelling banknote printing costs in a less-cash society, working paper, Vrije Universiteit Brussel, in preparation, 2013.